DOCTOR OF PHILOSOPHY

Structure- based inhibitor design for key enzymes of Tryposoma brucei

Striker, Waldemar

Award date:
2014

DOCTOR OF PHILOSOPHY

Structure- based inhibitor design for key enzymes of Tryposoma brucei

Waldemar Striker

2014

University of Dundee

UNIVERSITY OF DUNDEE

COLLEGE OF LIFE SCIENCES

STRUCTURE-BASED INHIBITOR DESIGN FOR KEY ENZYMES OF *TRYPANOSOMA BRUCEI*

WALDEMAR STRIKER

A THESIS SUBMITTED FOR THE DEGREE OF:

DOCTOR OF PHILOSOPHY
UNIVERSITY OF DUNDEE

29. January 2014

Declarations

I declare that the following thesis is based on the results of investigations conducted by myself, and that this thesis is of my own composition. Work other than my own is clearly indicated in the text by reference to the relevant researchers or to their publications. This dissertation has not in whole, or part, been previously submitted for a higher degree.

Waldemar Striker

I certify that Waldemar Striker has spent the equivalent of at least nine terms in the research work at the College of Life Sciences, University of Dundee, and that he has fulfilled the conditions of the Ordinance General No. 14 of the University of Dundee and is qualified to submit the accompanying thesis in application for the degree of Doctor of Philosophy.

Dr. Ruth Brenk

Acknowledgments

First I want to thank my supervisor Ruth Brenk for giving me the opportunity to do my PhD under her supervision, guidance and never giving up on me.

I would like to thank the whole Brenk-Group for their support and wonderful atmosphere in the office.

Many thanks to Ian Gilbert for his constructive suggestions and corrections of this thesis.

I want to thank all the people who were involved in the projects of this thesis: Raffaella Grimaldi, David Gray, Iva Hopkins Navratilova, Sabine Kuettel, Michael Ferguson and Daniel Fontaine.

A great support was David Robinson, who introduced me in crystallography, crystal structure-determination and NMR – screening. Thanks a lot!

Bill Hunter deserves my deeply gratitude to always be there for me and guide me through difficult times! I will never forget that!

My parents I want to thank for their support over all these years of studies.

Especially I want to thank Manuel Banzhaf for his major corrections of this thesis and for being a loyal and true friend for so many years. Who really needs more friends if he has already found the best one??

This thesis is dedicated to my wife Milana and my little family. Milana you made all these years of this PhD bearable and I want to thank you for your love and support in every situation.

Table of contents

List of figures

List of tables

List of equations

Abstract

Human African trypanosomiasis (HAT) also known as sleeping sickness is caused by a subspecies of *Trypanosoma brucei.* These parasites are transmitted by tsetse flies and endanger over 60 million people in Sub-Saharan Africa. Untreated, sleeping sickness is fatal, causing at least 48.000 deaths per year. Its treatment remains complicated since the currently available drugs show high toxicity and are too expensive to be ubiquitously distributed in the affected third world economies. Additionally, emerging drug resistance towards the most clinically relevant anti HAT drugs, drastically limits treatment options and makes it imperative to conduct research to find safer and more efficient drugs to treat this terrible disease.

This thesis describes the hit identification and hit validation for two validated targets for HAT: *Tb*6PGDH and *Tb*UGP. For hit identification different techniques like *in silico* virtual screening, NMR lead-like fragment screening and HTS were used. For *Tb*UGP a very first drug-like, competitive inhibitor with a pIC_{50} of 3.53±0.04 and a Hill slope of 1.1±0.1 was discovered. Additionally this thesis describes the determination and validation of the *in silico* proposed binding mode using mutation studies and crystallisation techniques

1 Introduction

1.1 Drug discovery and drug design

1.1.1 Drug discovery process

In 7000 years of development of mankind, drug discovery has always played a crucial role (Lindesmith 1968). The more advanced the medicinal knowledge of a community was, the healthier and therefore longer-living and historically influential people were. Although hundreds of years ago many drugs were known, which were isolated from plants, animals and minerals, the knowledge and discovery of new drugs developed very slowly. With the beginning of the modern age, drugs from all over the world became available for trade with the consequence that drug discovery gathered speed. At the end of the 19th century organic-synthetically drugs had a major success. Drugs like sulfonamide (Prodrug Prontosil discovered 1932 by Gerhard Domagk) (Bickel 1988; Schirren 1988; van Miert 1994) or aspirin (isolated in 1897 by Felix Hoffmann)(Sneader 2000; Cheng 2007; Folts 2007; Schror 2009; Wick 2012) are still used ubiquitous. Nowadays, the drug discovery process is an established multibillion dollar project involving thousands of people worldwide. Modern drug discovery must meet many challenges like research and development speed, cost and quality, leadership and management, selecting the right

pharmacologic target and the right chemical lead (Elebring, *et al.,* 2012). The current strategy starts with the search for a suitable target (Figure1.1).

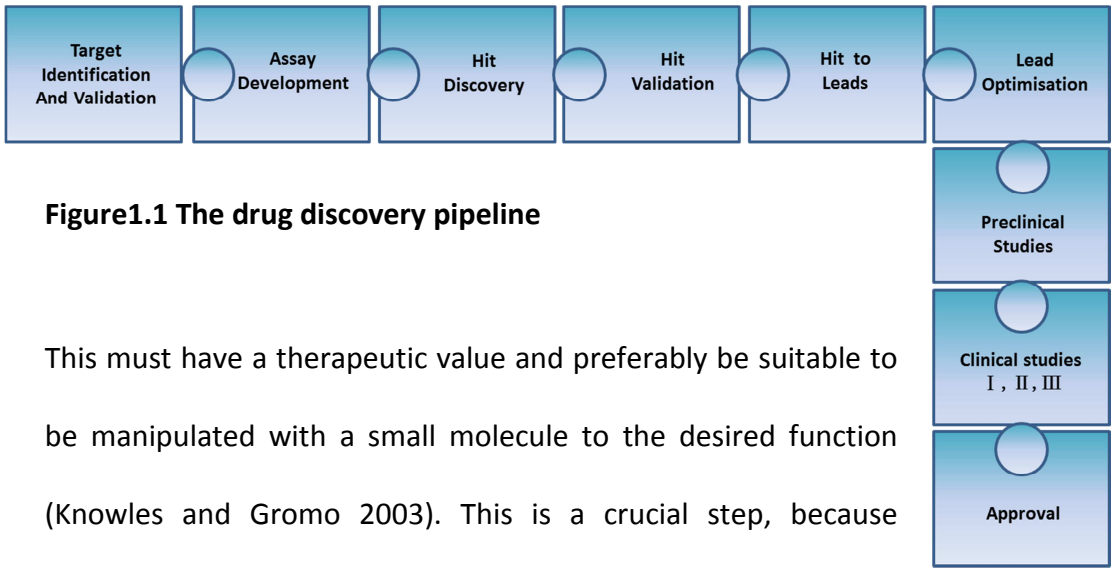


Figure1.1 The drug discovery pipeline

This must have a therapeutic value and preferably be suitable to be manipulated with a small molecule to the desired function (Knowles and Gromo 2003). This is a crucial step, because concentrating on the wrong target is very time and cost intensive. For the selected target an assay must be developed which delivers robust and reproducible results. This assay may then be used in the hit discovery step. There are two possible ways to find hits: *in silico* screening and experimental screening. The *in silico* screening method will be described in detail later (1.1.3).

Nuclear magnetic resonance (NMR) and surface plasmon resonance (SPR) are suitable for the hit validation step. Hereby, all hits must be retested to filter out false positives and confirm the other hits. False positive compounds mostly absorb at the wavelength where the assay is carried out, are fluorescent, interfere with the assay setup or are chemically impure and reactive. For this reasons the compounds are tested for purity using LCMS or NMR and are either repurchased or resynthesized for the retest. Usually a dose response curve will be measured in a

biochemical or cell assay to identify the binding mode of the compound (aggregator, allosteric inhibitor, etc.) and to rank the hits by their potency (IC_{50}, EC_{50}) (Alphey*, et al.,* 2012). In the next step, the initial hits need to be improved to leads. Therefore, promising compounds are chemically modified and further tested. Here, not only the potency of the compound can be improved but also other important properties like lipophilicity, solubility, size, ligand efficiency and selectivity. All these modifications help to establish the structure activity relationship (SAR) of the preferred compound series. Subsequently, the candidate reaches the lead optimisation step, where it is further improved to meet the required pharmacokinetic and pharmacodynamics profile. In the preclinical studies the compounds are tested *in vivo* and *in vitro* for toxicity. Only if the chemical passes this tests, it will be used for humans in the clinical studies Ⅰ, Ⅱ and Ⅲ. In the very last step, all the studies must be provided to the relevant agency in order to get approval as a new drug, e.g. the *Medicines and Healthcare products Regulatory Agency* (MHRA) for a UK licence or the *European Medicines Agency* (EMA) for a Europe wide licence.

1.1.2 Experimental screening

In the experimental screening methods like high throughput screening (HTS), nuclear magnetic resonance (NMR), mass spectrometry (MS) or surface plasmon

resonance (SPR) are used to screen compound libraries for activity against the target.

HTS plays a dominant role in modern drug discovery. The aim is to find active compounds (hits) by screening a large database of chemically diverse compounds against proteins, cells or other targets. This is possible, because HTS consists of multiple automated steps like liquid dispensing, compound transfer and signal capturing. This automation leads to low systematic failures during the screen and can generate highly reproducible and reliable data (Shun*, et al.,* 2011). Unfortunately, HTS often produces also many false positive hits that slow down the process of drug finding (Crisman*, et al.,* 2007; Posner*, et al.,* 2009; Sink*, et al.,* 2010; Bocker*, et al.,* 2011; Liu*, et al.,* 2012; Prummer 2012). Screening by NMR has been established as an alternative method (Bhunia*, et al.,* 2012; Jordan*, et al.,* 2012; Mizukoshi*, et al.,* 2012; Stark and Powers 2012; Wirmer-Bartoschek and Bartoschek 2012). Compared to HTS, it provides a cheap method to detect even weak protein-ligand interactions, dissociation constant, identify a ligand binding site and generate a complex structure (Stark and Powers 2012).

Since Lipinski *et al.,* published the rule of 5 where he discovered, that orally available drugs do not violate more than the following criteria (Lipinski*, et al.,* 1997; 2001):

- Molecular weight (MW) < 500 Daltons
- Logarithmic octanol/water partition coefficient (logP) < 5

- Hydrogen-bond donors ≤ 5
- Hydrogen-bond acceptors ≤ 10

the compound databases could be easily filtered applying these rules (Matter*, et al.,* 2001). Shortly after Lipinski, Hann *et al.,* discovered, that drug-like molecules were still too complex (too high MW, too many heavy atoms, etc.) to be used as a good starting point for drug discovery (Hann*, et al.,* 2001) and suggested to use lead-like compounds instead. These lead-like compounds are less complex (less MW, less number of rings and rotatable bonds) and less hydrophobic (lower CLogP)(Oprea*, et al.,* 2001). The advantage in using lead-like compounds as a starting point for drug discovery is, that these small, active molecules can gain in MW and CLogP while being optimised to become a drug-like candidate. Recently, the screening of fragment-like libraries was established (Badger 2012; Bower and Pannifer 2012; Duong-Thi*, et al.,* 2012; Kumar*, et al.,* 2012). Keeping this in mind, the NMR screen is sensitive enough to detect hits from small, fragment-like libraries. With the low MW of the fragments, the chemical space can be exploited much more effective. These fragment-like compounds are smaller (usually < 18 heavy atoms) and the hits therefore usually less potent. Traditional screening methods such as biochemical assays are often not effective enough to detect such weak binding. To detect weakly bound fragment-like compounds, biophysical screens like X-ray crystallography or NMR are used. The big advantage of X-ray crystallography is that it additionally gives the binding mode of the fragment.

1.1.3 *In silico* (virtual) screening

Modern drug discovery would not be possible without using computational methods. These can be applied for hit-discovery using virtual screening of small molecule libraries, hit to lead optimisation and also to improve the psychochemical properties in the lead optimisation (Bernardo and Tong 2012; Cheng*, et al.,* 2012; Ma*, et al.,* 2012). There are several advantages of *in silico* screening compared to experimental screening methods. It is both, much cheaper and much faster compared to any experimental screening approach. Using computational power, the chemical space can be exploited more effectively by screening a large number of molecules. Even for a small compound with 12 heavy atoms, the number of potential drug-like molecules (not including 3- and 4-membered ring structures) has been estimated to be 10^7 (Fink, et al., 2005). This number is raising to 10^{60} considering a compound with 30 heavy atoms (Bohacek*, et al.,* 1996). Considering this large number, the only way to manage, screen and evaluate such datasets is by using *in silico* methods. The two major approaches of *in silico* drug screening are structure-based and ligand-based screening (Figure 1.2). Ligand-based screening is usually used if a set of structurally diverse ligands is available. The assumption is, that compounds found with a similarity search, will have a similar activity as the lead compound (Sheridan and Kearsley 2002; Bender*, et al.,* 2009; Vilar*, et al.,* 2012). The method used in this thesis was a structure-based virtual screening (SBVS). This method is applied when a 3D structure of the biological target or a

homology model is available. When a binding site/pocket of the target is known, libraries of small molecules can be screened and the compounds will be predicted to bind into this site/pocket (Hurko 2012) using molecular docking programs like Gold (Jones, *et al.,* 1997), DOCK (Lorber and Shoichet 1998) , Glide (Friesner, *et al.,* 2004; Halgren, *et al.,* 2004), FlexX (Kramer, *et al.,* 1999), Fred (McGann, *et al.,* 2003) and LigandFit (Venkatachalam, *et al.,* 2003).

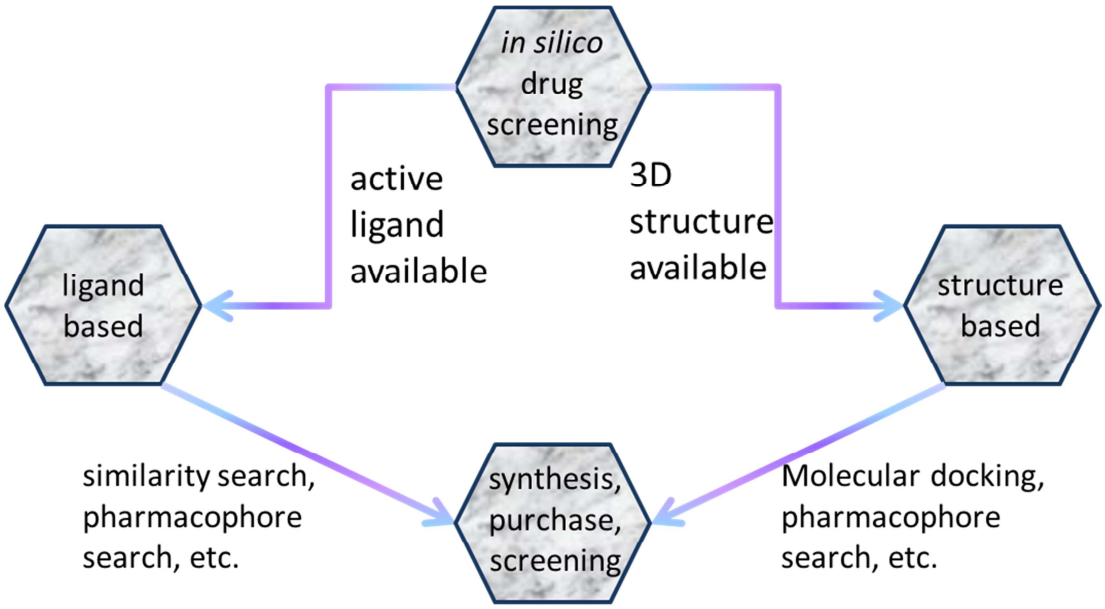


Figure 1.2 *In silico* drug screening flowchart

How strong a compound binds into the active site is dependent on what kind of interactions the compound forms with the protein, how many interactions it makes, its desolvation energy and ultimately what the free energy of binding is. This is predicted using a mathematical algorithm, also called the "scoring function". Each docked compound can be ranked and compared by the docking score (Coupez and Lewis 2006). Each docking program has its own calculation of the score (Feher and

Williams 2010; Pencheva*, et al.,* 2010; Abreu*, et al.,* 2012). The hits with the best score will then be either synthesized or bought for biological testing. The same principle can be used for the lead optimization process. The modified lead compounds are docked into the active site and then scored and ranked. Only the modification which gave a higher score will then be tested in a biological screen.

The disadvantages of *in silico* screening are certainly the many false positives and negatives (Ekins*, et al.,* 2007; 2007). It is still not possible to predict protein flexibility, molecule conformation and promiscuity perfectly due to the accurate calculation of the binding energy. The other problem is, that not all molecules, that could be found by *in silico* screening are available for testing. For this reason mostly only compounds, which are already commercially available are screened.

1.1.3.1 Docking program DOCK3.5.54

In this thesis mainly the docking program DOCK3.5.54 (Lorber and Shoichet 2005) was used for SBVS. It is a variation of DOCK, which was initially created in the 1980s by the Kuntz group and was the first docking program available (Kuntz*, et al.,* 1982).

Placing a ligand into an enzyme is a challenging task considering the degree of freedom of the ligand and the enzyme. Even when the ligand is considered to be rigid, it has already a degree of freedom of 6. And if the flexibility of ligand and protein is considered additionally with every rotated bond, the degree of freedom

increases exponentially, so therefore the systematic exploitation of all possible binding modes is computationally not feasible for a docking review.

The docking program would have to be able to dock all possible orientations and conformations of the ligand, considering translation and rotation. To simplify this task, DOCK3.5.54 considers the protein as static. The pre-generated conformers (for example using OMEGA (Hawkins*, et al.,* 2010)) are then superimposed on a common ring fragment (usually an aromatic ring system)(Lorber and Shoichet 1998). Next, matching spheres are placed into the binding site of the protein. Subsequently, DOCK3.5.54 places the fragments onto the sphere centre, so that the fragment atoms are as close as possible to the sphere centres. Together they are used to generate a translational-rotational matrix which allows the fragment to be oriented in the binding site and with it the pre-calculated conformers.

DOCK3.5.54 uses a force-field-based scoring function (Equation 1.1), which contains terms of electrostatic Energy (E_{elec}), van der Waals (E_{vdw}) and the correction for ligand desolvation energy (ΔG_{desolv}) (Wei*, et al.,* 2002; Graves*, et al.,* 2005; Lorber and Shoichet 2005).

$$E = E_{elec} + E_{vdw} + \Delta G_{desolv}$$

Equation 1.1 Scoring function used in DOCK3.5.54

1.2 Human African trypanosomiasis (HAT)

Human African trypanosomiasis (HAT), also known as sleeping sickness, is a pandemic parasitic disease found in the sub-Saharan Africa where it threatens millions of people in 36 countries. This fatal disease belongs to one of the most neglected diseases and is mostly prevalent in rural areas (Balasegaram*, et al.,* 2008; Simarro*, et al.,* 2011). Its prevalence has changed in the last 100 years. In the 1960s, transmission was practically interrupted in all endemic areas because of control and intervention programs (Brun*, et al.,* 2010). The rarity of cases led to lower interest in surveillance and caused the disease to re-emerge in 1980s (Simarro*, et al.,* 2011). For this reason a non-profit drug research and development organization DNDi (Drugs for Neglected Diseases Initiative) was established (Balasegaram*, et al.,* 2008; Chatelain and Ioset 2011) to develop new treatments for the most neglected diseases like leishmaniasis, chagas disease, malaria and HAT.

HAT is transmitted by the tsetse fly (*Glossina* spec.) and is caused by two subspecies of *Trypanosoma brucei* (*T. brucei or Tb*), the *T. brucei gambiense* (West Africa) and *T. brucei rhodesiense* (east Africa) (Fevre*, et al.,* 2008). The symptoms of HAT depend on the subspecies and the stage of infection. Initially it starts with fever, headache, joint pains and itching and later listlessness, disordered sleep and neuromuscular dysfunction (Stich*, et al.,* 2002; Checchi and Barrett 2008; Courtin*, et al.,* 2008; Fevre*, et al.,* 2008; Brun*, et al.,* 2010; Malvy and Chappuis 2011). There are two stages of infection, firstly the haemolymphatic stage, where *T. brucei* enters

the blood stream and the lymph system, where it replicates. The second stage is the meningo-encephalitic stage, where the parasite invades the central nervous system. In addition, *T. brucei gambiense* is characterized by a chronic progression with an average duration of around 3 years and can be mistaken for a chronic haemopathy condition (Malvy and Chappuis 2011). *T. brucei rhodesience* disease on the other hand presents usually in an acute illness and leading to death within months. Most non-endemic cases of HAT are *T brucei rhodesience* diseases (Migchelsen*, et al.,* 2011; Simarro*, et al.,* 2012). With 94 reported cases from 2000 to 2010 outside the endemic areas, HAT represents also a risk for travellers and migrants (Simarro*, et al.,* 2012).

Untreated HAT always leads to coma and death. There are four main drugs, which are currently used for HAT treatment (Figure 1.3). All of them have significant limitations due to toxicity, administration and treatment regimes. No new chemical drug has been approved since eflornithine in 1990 (Jacobs*, et al.,* 2011). The only breakthrough so far was a combination treatment with nifurtomox/eflornithine (NECT) which was less toxic than eflornithine alone and easier to administer. Pentamidine and suramin are available for early-stage *T. brucei gambiense* disease. None of the drugs can cross the blood-brain barrier and are therefore useless for second stage treatment. Pentamidine reduces the mitochondrial membrane potential and binds to nucleic acids. It must be given intramuscularly and can cause hypotension (low blood pressure), hypoglycemia (diminished levels of glucose in blood), leukopenia (decreased number of white blood cells), hepatitis (liver

inflammation), nephrotoxicity (kidney poisoning) and pancreatitis (inflammation of the pancreas) (Kappagoda*, et al.,* 2011). Suramin inhibits multiple trypanosome metabolic enzymes. Due to its toxicity it is only given as a second-line treatment. Suramin can cause exfoliative dermatitis (erythema and scaling of the skin), neuropathy (damage to nerves) and fatal hypersensitivity reaction (body reacts with an exaggerated immune response). For the second stage only eflornithine and melarsoprol are used. Eflornithine inhibits ornithine decarboxylase. It is less toxic than melarsoprol but also less reliable against *T. brucei rhodesiense*. The side effects could be fever, rash, peripheral neuropathy and diarrhoea. The mostly used drug against the second HAT stage is melarsoprol, despite its highest toxicity. Melarsoprol causes in 5 – 10% of the patients an encephalopathy which is in half of the cases fatal. It also causes fever, thrombocytopenia, abdominal pain and vomiting.

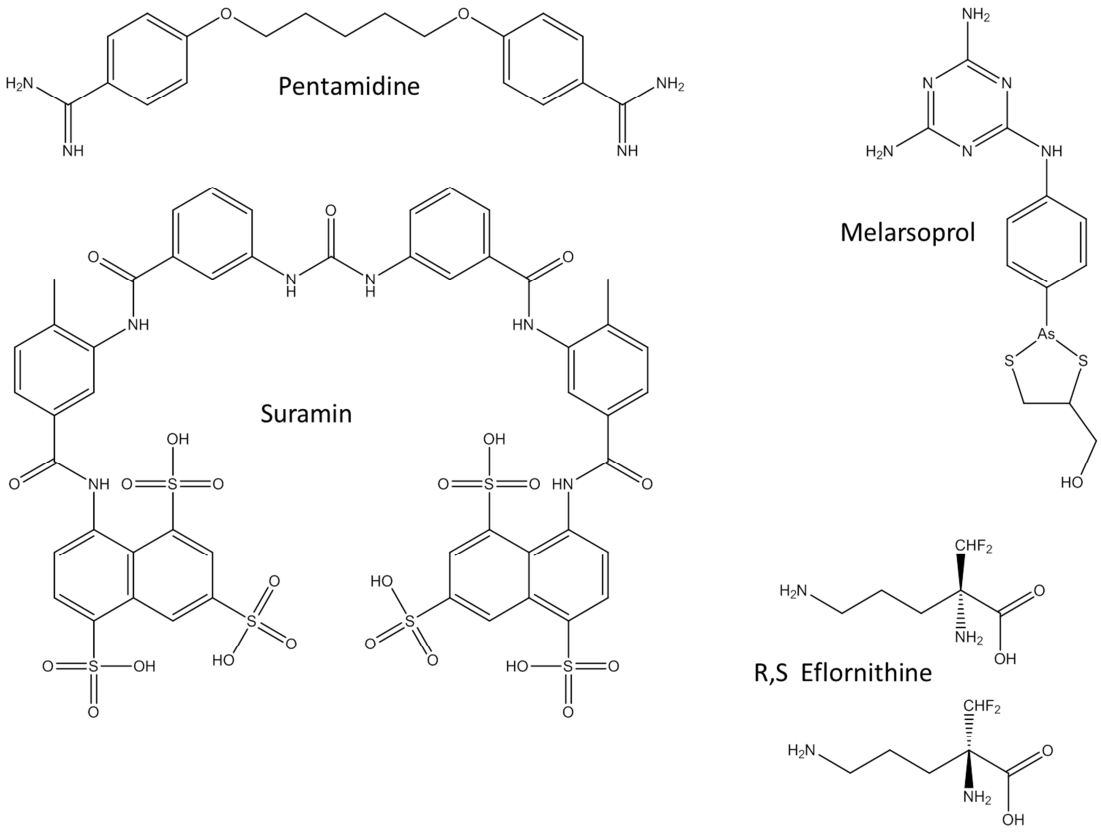


Figure 1.3 Drugs currently used for HAT treatment

The insufficiency of the current drugs and treatment of HAT shows clearly, that new, safer and more effective drugs are badly needed.

1.3 *T. brucei* 6-phosphogluconate dehydrogenase (*Tb*6PGDH) as a target for HAT

The pentose phosphate pathway (PPP) (Figure 1.4 oxidative part) is present in most species (also humans) and is another key pathway of glucose metabolism (Duschak 2011; Leroux*, et al.,* 2011; Maugeri*, et al.,* 2011; Stern*, et al.,* 2011). The enzyme

6PGDH catalyse the conversion of 6-Phosphogluconate (6PG) to Ribulose-5-phosphate (Ru5P) by generating NADPH and CO_2.

Using a drug target that is present in parasite and mammalian cells seems at first glance not to be a good choice. Nevertheless, it has been demonstrated in case of the key glycolytic enzyme glyceraldehyde-3-phosphatedehydrogenase (GAPDH) that differences in the binding site of coenzyme NAD^+ are sufficient to selectively inhibit trypanosomatid enzymes (Aronov*, et al.,* 1999).

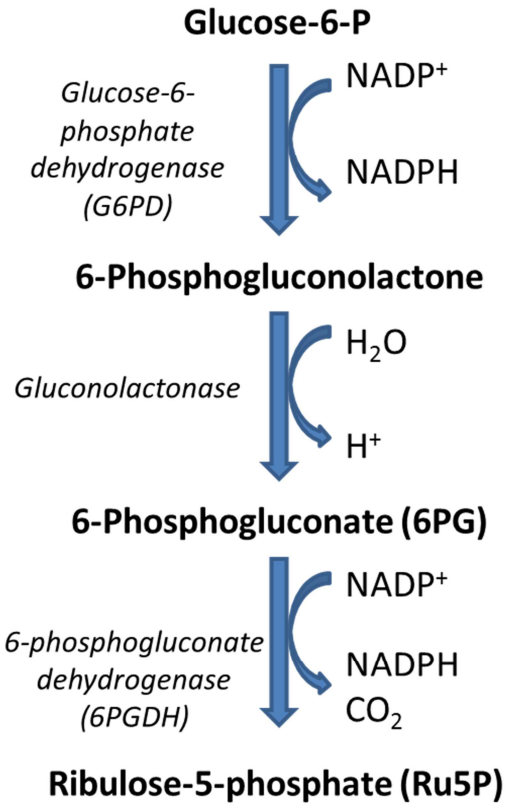


Figure 1.4 Oxidative phase of the pentose phosphate pathway
In italic are the enzymes involved in the reaction and in bold the sugar educts and products.

Two $NADP^+$ dependent enzymes of PPP, glucose-6-phosphate dehydrogenase (G6PD) and 6-phosphogluconate dehydrogenase (EC 1.1.1.44) have a major function in most organisms. They generate the reduced coenzyme NADPH which protects the organism against oxidative stress and is essential for a variety of reductive biosynthetic reactions, such as lipid production. A G6PD deficiency in humans causes haemolytic anaemia, an abnormal breakdown of red blood cells. The PPP was also determined to be present in both cultured procyclic and bloodstream forms of *T. brucei* (Cronin*, et al.,* 1989). Sequence analysis has shown, that the enzyme 6PGDH had an unusual evolution in *T. brucei* and was only distantly related to the one in mammals (Barrett and Le Page 1993). The differences were evident by the fact, that trypanocidal drugs melarsoprol, cymelarsan and suramin were more potent against *Tb6PGDH* than against the mammalian enzyme (Hanau*, et al.,* 1996). The enzyme *Tb*6PGDH was validated as a drug target using RNA interference technology (Bastin*, et al.,* 1998; Craig*, et al.,* 1998; Fire*, et al.,* 1998) to switch off the gene encoding 6PGDH (Hanau*, et al.,* 2004). This led to an accumulation of the substrate 6-phosphogluconate (6PG) which itself is an inhibitor of the key glycolytic enzyme phophoglucose isomerase in the second step of the Embden-Meyerhof glycolytic pathway (Rovere and Gastaldi 1967; Marchand*, et al.,* 1989). As a consequence more 6-phosphate enters the PPP and increases the production of 6PG even more. This positive feedback loop is fatal for *T. brucei*, because of its dependence on the glycolysis pathway for energy production. Furthermore, the parasite would not produce enough NADPH to protect itself against oxidative stress

and to generate carbohydrate intermediates used in nucleotide and other biosynthetic pathways. *Tb*6PGDH was chemically validated using analogues of high-energy intermediates and transition-state analogues. Both were able to selectively inhibit this drug target (Dardonville*, et al.,* 2004) (Figure 1.5).

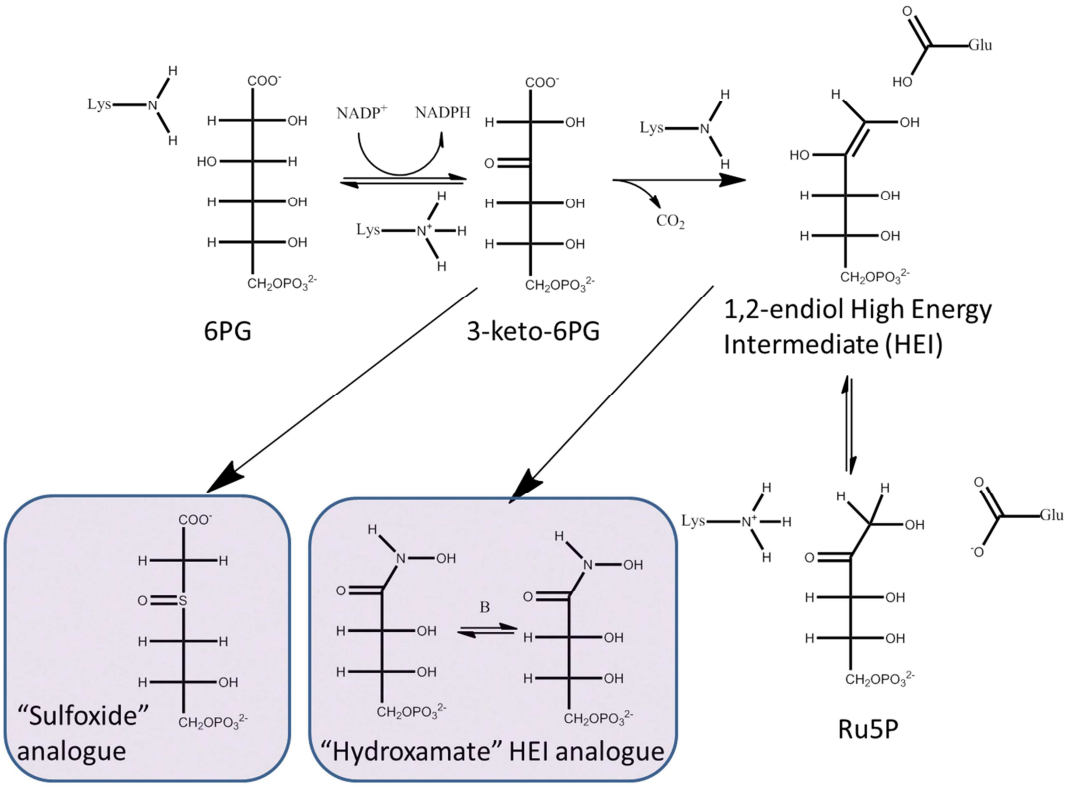


Figure 1.5 Mechanism of 6PGDH reaction and intermediate analogues
Mechanism was adopted from (Wang and Li 2006). Lysine serves as a base and glutamine as acid in this reaction mechanism (Karsten*, et al.,* 1998).

A number selective inhibitors for *Tb*6PGDH were found and optimized (Bertelli*, et al.,* 2001; Pasti*, et al.,* 2003; Ruda*, et al.,* 2007; Ruda*, et al.,* 2010). A virtual fragment screening which addressed the phosphate binding site of 6PGDH identified several fragments with high ligand efficiencies and IC_{50} values in the low micro molar range (Ruda*, et al.,* 2010). However, the binding modes of the

compounds have not been determined. Therefore, it is unclear how they bind to the enzyme, hindering their optimization. The crystal structure of 6PGDH is currently known for fourteen different organisms including *Tb* (Phillips*, et al.,* 1998), *Geobacillus stearothermophilus* (*Gs*) (Cameron*, et al.,* 2009), *Lactococcus lactis* (Sundaramoorthy*, et al.,* 2007) and recently from *homo sapiens* (PDB 2JKV)(to be published).

1.4 *T. brucei* UDP-glucose pyrophosphorylase (*Tb*UGP) as a target for HAT

It was shown, that glycoproteins are crucial for the survival, infectivity and *de novo* biosynthesis of sugar nucleotides of the parasites (Turnock and Ferguson 2007). The sugar nucleotides uridine-diphosphate galactose (UDP-Gal) (Roper*, et al.,* 2002; Roper*, et al.,* 2005; Urbaniak*, et al.,* 2006), UDP-N-acetylglucosamine (UDP-GlcNAc)(Stokes*, et al.,* 2008) and GDP-fucose (GDP-Fuc) (Turnock*, et al.,* 2007) have been shown by mutation studies to be essential for parasites growth. The nucleotides are the primary source of sugar for most glycosylation reactions. They can be formed either in the salvage pathway, by recycling existing sugar/sugar nucleotides that were formed during degradation of RNA/DNA or in *de novo* by biosynthesis of complex molecules from simple molecules like sugar or amino acids.

The sugar nucleotide UDP-Glc (Figure 1.6) is the donor of glucose in many different pathways and in *Tb* crucial in the synthesis of several glucose containing glycolipids, glycoproteins and a variety of secondary metabolites (Flores-Diaz*, et al.,* 1997).

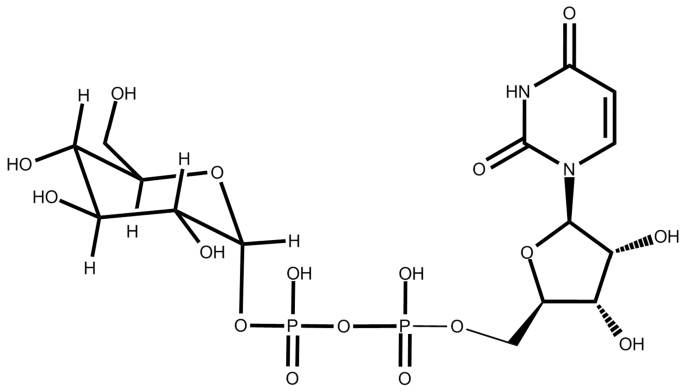


Figure 1.6 Chemical structure of UDP-Glc

Furthermore it plays an important role for the “quality control” of newly synthesized glycoproteins in the endoplasmic reticulum (Hammond and Helenius 1995).

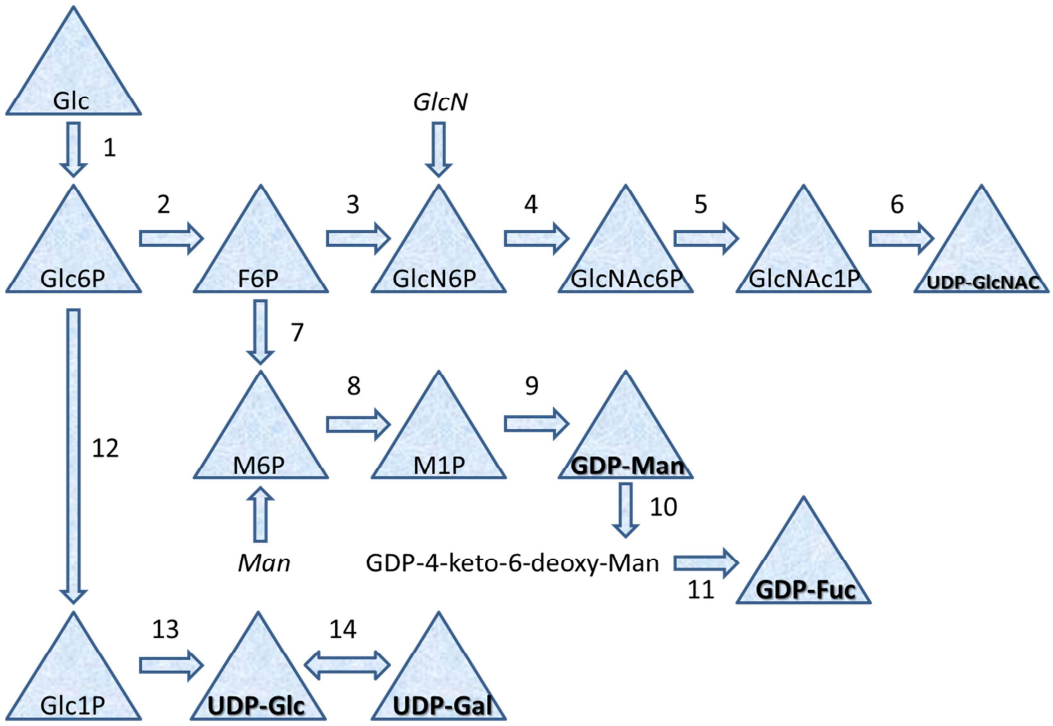


Figure 1.7 Summary of the sugar nucleotide biosynthesis pathway in *T. brucei*
Sugar nucleotide biosynthesis pathway for Tb adopted from (Turnock and Ferguson 2007). The sugar nucleotides used for glycoconjugate biosynthesis in bold. Salvage pathways are in italics. Glc = Glucose; Glc6P = Glucose-6-Phosphate; F6P = Fructose-6-Phosphate; GlcN6P = Glucoseamine-6-Phosphate; GlcN = Glutamine; GlcNAc6P = N-Acetyl-Glucosamine 6-Phosphate; GlcNAc1P = N-Acetyl-Glucosamine 1-Phosphate; UDP-GlcNAc = UDP-N-acetyl-Glucosamine; M6P = Mannose-6-Phosphate; M1P = Mannose-1-Phosphate; GDP-Man = Guanosine Diphosphate Mannose; GDP-Fuc = Guanosine Diphosphate Fucose; Glc1P = Glucose-1-Phosphate; UDP-Glc = UDP-Glucose; UDP-Gal = UDP-Galactose. The numbers stand for the following enzymes: 1 = Hexokinase; 2 = Glucose-6-phosphate isomerase; 3 = Glucosamine-fructose-6-phosphate aminotransferase; 4 = Glucosamine-phosphate N-acetyltransferase; 5 = Phosphoacetylglucosamine mutase, 6 = UDP-N-acetylglucosamine pyrophosphorylase; 7 = Phosphomannose isomerase; 8 = Phosphomannomutase; 9 = Mannose-1-phosphate guanyltransferase; 10 = GDP-mannose 4,6-dehydratase; 11 = GDP-L-fucose synthetase; 12 = Phosphoglucomutase; 13 = UDP–glucose pyrophosphorylase; 14 = UDP-galactose 4-epimerase.

Moreover, UDP-Glc is presumably a donor for the modified DNA base β-D-glucosyl-hydroxymethyluracil, called J. The base J has been found in *Tb* in the 70-basepair

repeats and variant surface glycoprotein (VSG) gene of the telomeric FSG gene expression sites (van Leeuwen*, et al.,* 1997; van Leeuwen*, et al.,* 1998). This VSG is a protective coat of $5x10^6$ GPI-anchored homodimers that helps the parasite to undergo the immune attack of the host by replacing the VSG coat by antigenetically different VSG molecules (Borst*, et al.,* 1996; Cross 1996). As shown by Urbaniak, Turnock *et al.,* (2006) knockout studies, the galactose metabolism is essential for the survival of *Tb.* The only way for *Tb* to synthesise galactose is by epimerisation of UDP-Glc to UDP-Gal by UDP-galactose 4-epimerase (Figure 1.7, enzyme 14). By interrupting the production of UDP-Glc by inhibiting *Tb*UGP (EC 2.7.7.9), the parasite dies after 96 hours.

Biosynthesis of UDP-Glc in *Tb* follows an ordered mechanism (Figure 1.8). First UTP binds to *Tb*UGP and is stabilised by Mg^{2+} followed by the binding of Glc-1P and the reaction into UDP-Glc and pyrophosphate (PP_i). This was confirmed by SPR, where the binding of Glc-1P could only be measured, when UTP was added into the running buffer (Hopkins, Navratilova *et al.,* unpublished results).

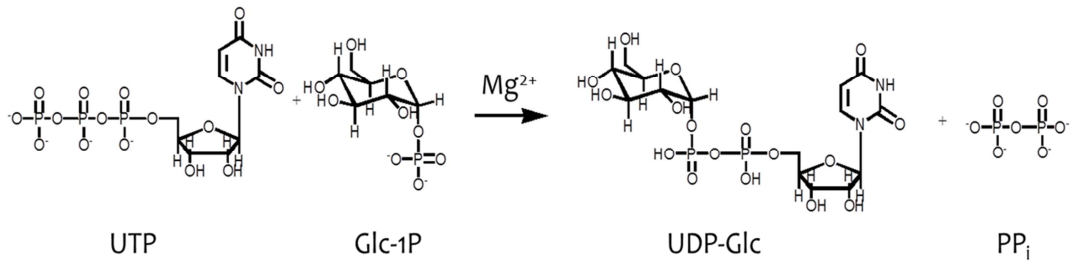


Figure 1.8 Ordered reaction mechanism of *Tb*UGP

Unfortunately still no drug-like inhibitors are known for *Tb*UGP. The only inhibitor known for *Tb*UGP is the non- hydrolysable UTP-analogue UTP-α-S (compound **7** from Table 3.5). However, this compound is not drug-like.

1.5 Objectives

1.5.1 Structure-based hit discovery for the key enzymes *Tb*6PGDH and *Tb*UGP

The aim of this thesis was it to use *in silico* and experimental methods to identify novel, active, fragment-like compounds for two validated targets, *Tb*6PGDH and *Tb*UGP (1.3 and 1.4), against HAT. To achieve this, fragment-libraries were screened for hit compounds using NMR. Further, structure-based virtual screening was used to screen a large library of commercially available compounds for *Tb*UGP inhibition. To cover more chemical space, hit-discovery for *Tb*UGP was complemented by a biochemical screen of a lead-like library with over 72 000 compounds.

1.5.2 Hit validation, SAR and evaluation of binding mode of discovered hits

Potential hits, discovered by *in silico* screening and experimental screening efforts, were further validated using biophysical methods such as SPR and NMR. These methods were used in order to characterize potency, competitive binding and selectivity of the preselected hits. Furthermore, the aim was to unravel the binding mode of the inhibitors and SAR using mutation studies and crystallography.

2 Hit discovery for 6PGDH

2.1 Material and Methods

All chemicals and reagents where purchased from Sigma-Aldrich, unless otherwise stated. Virtual screening compound **1** was purchased from Chembridge, **2** from LABOTEST and **3** from Enamine.

2.1.1 Overexpression and purification of *Gs*6PGDH

Table 2.1 Sequence of His-tagged *Gs*6PGDH

HHHHHHSSGLEVLFQGPGHMAKHQIGVIGLAVMGKNLALNIESKGYSVAVYNRLREKTDEFL
QEAKGKNIVGTYSIEEFVNALEKPRKILLMVKAGAPTDATIEQLKPHLEKGDIVIDGGNTYFKDT
QRRNKELAELGIHFIGTGVSGGEEGALKGPSIMPGGQKEAHELVRPIFEAIAAKVDGEPCTTYIG
PDGAGHYVKMVHNGIEYGDMQLIAEAYFLLKHVLGMDAAELHEVFADWNKGELNSYLIEITA
DIFTKIDEETGKPLVDVILDKAGQKGTGKWTSQNALDLGVPLPIITESVFARFLSAMKDERVKAS
KVLAGPAVKPFEGDRAHFIEAVRRALYMSKICSYAQGFAQMKAASEEYNWNLRYGDIAMIFR
GGCIIRAQFLQKIKEAYDRDPALSNLLLDSYFKDIVERYQDALREIVATAAMRGIPVPGSASALAY
YDSYRTAVLPANLIQAQRDYFGAHTYERVTKKAIPHTEWLK

Table 2.2 Biochemical properties of His-tagged *Gs*6PGDH

Number of amino acids: 488

Molecular weight: 54017.9

Theoretical pI: 6.44

Abs 0.1% (=1 g/l): 0.989 at 280 nm (Artimo*, et al.,* 2012)

Table 2.3 Sequence of cleaved *Gs*6PGDH (crystallization construct)

IESKGYSVAVYNRLREKTDEFLQEAKGKNIVGTYSIEEFVNALEKPRKILLMVKAGAPTDATIEQL

KPHLEKGDIVIDGGNTYFKDTQRRNKELAELGIHFIGTGVSGGEEGALKGPSIMPGGQKEAHEL

VRPIFEAIAAKVDGEPCTTYIGPDGAGHYVKMVHNGIEYGDMQLIAEAYFLLKHVLGMDAAEL

HEVFADWNKGELNSYLIEITADIFTKIDEETGKPLVDVILDKAGQKGTGKWTSQNALDLGVPLPI

ITESVFARFLSAMKDERVKASKVLAGPAVKPFEGDRAHFIEAVRRALYMSKICSYAQGFAQMK

AASEEYNWNLRYGDIAMIFRGGCIIRAQFLQKIKEAYDRDPALSNLLLDSYFKDIVERYQDALREI

VATAAMRGIPVPGSASALAYYDSYRTAVLPANLIQAQRDYFGAHTYERVTKKAIPHTEWLK

Table 2.4 Biochemical properties of cleaved *Gs*6PGDH

Number of amino acids: 448

Molecular weight: 49724.9

Theoretical pI: 6.01

Abs 0.1% (=1 g/l): 1.074 at 280 nm (Artimo*, et al.,* 2012)

Procedure

*Gs*6PGDH was overexpressed and purified as described by (Cameron*, et al.,* 2009). In brief, a *Gs*6PGDH full length ORF, cloned into a modified pET15b vector with N-terminal hexa-histidine tag coupled to a Tobacco Etch Virus (TEV) protease recognition site was used. The expression was performed in Bl21(DE3)pLysS cells. Therefore, a single colony was grown with constant shaking (200rpm) in auto-induction media (see Recipe for auto-induction media, page 115, appendix) supplemented with 50 mg/l carbenicillin, for approximately 2 h at 37 °C, followed by a second incubation step for 22 h at 22 °C. Cells were harvested by centrifugation (3500 g, 20 min, 277 K on BECKMAN J6-MC) and subsequently resuspended in resuspension buffer A (50 mM Tris-HCl pH 7.4, 250 mM NaCl, 50 µg Dnase, 50 µg lysozyme and EDTA-free protease inhibitor cocktail (Roche) according to manufactures instructions). In the next step the cells were disrupted by sonification and centrifuged (50,000 g, 30 min, 277 K on BECKMAN Avanti-J25) to separate the cell extract and the membrane pellet. The Supernatant was filtered and applied to a HisTrap HP 5 ml column using an ÄKTA-purifier.

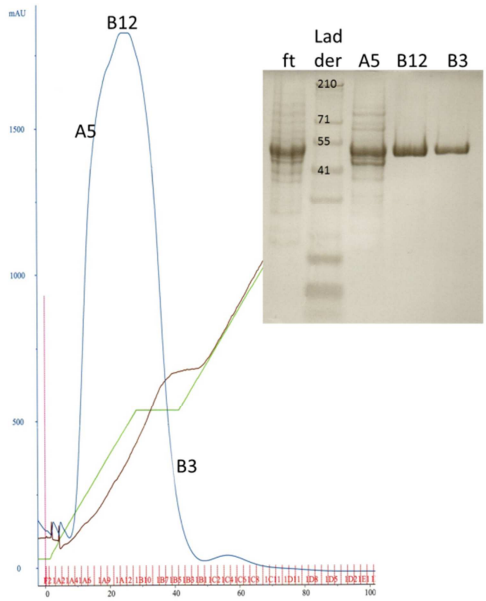


Figure 2.1 Chromatogram of *Gs*6PGDH on a HisTrap HP 5 ml column
The chromatogram displays the UV absorbance on the y-axis in blue in mAU, volume on the x-axis in ml, in green a gradient of imidazole in buffer A, in brown is the conductivity of the solution and in red the collection tubes. The SDS-page gel on the right shows the following from left to right: ft = flow through HisTrap with buffer A, protein ladder in kDa (SeeBlue), samples from tubes A5, B12, B3. Protein collected from A5-B3 for further purification.

Retained His-tagged protein was eluted by a gradient of imidazole starting from 0 to 1 M buffer B consisted of buffer A with 1 M imidazole. The eluted protein was cleaved for 2 h at 30 °C by adding 1 mg TEV protease (lab produced) per 15 mg 6PGDH and dialysed in buffer A overnight at 4 °C. Next, 6PGDH was applied to a His-Trap column to remove uncleaved protein.

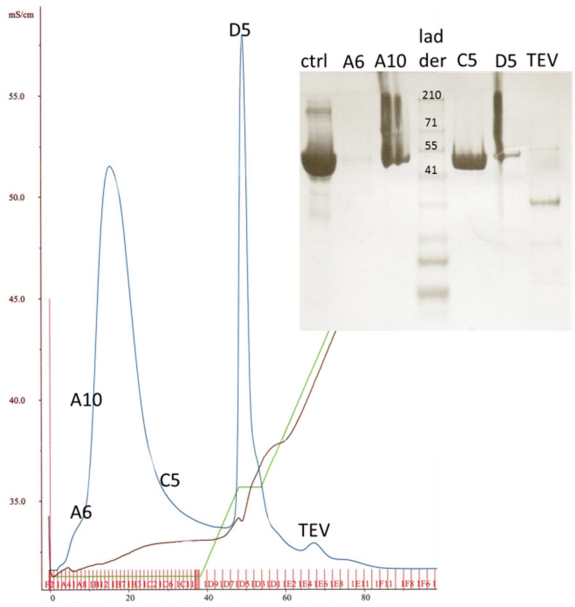


Figure 2.2 Chromatogram of *Gs*6PGDH on a HisTrap HP 5 ml column after TEV – cleavage

The chromatogram displays the UV absorbance in blue, in green a gradient of imidazole in buffer A, on the y-axis in brown is the conductivity of the solution in mS/cm and in red the collection tubes. Fractions that were analyzed by SDS-page are labeled in the chromatogram accordingly. The SDS-page gel on the right shows the following from left to right: ctrl = defrosted, TEV - cleaved and purified 6PGDH enzyme as a control; samples from tubes A6, A10, C5 which were eluated without imidazole; protein ladder in kDa (SeeBlue) in the middle; D5 = uncleaved protein; TEV = cleaved TEV - Tag

The elution containing cleaved protein in the resuspension buffer was subsequently purified using a Superdex 75 16/60 column.

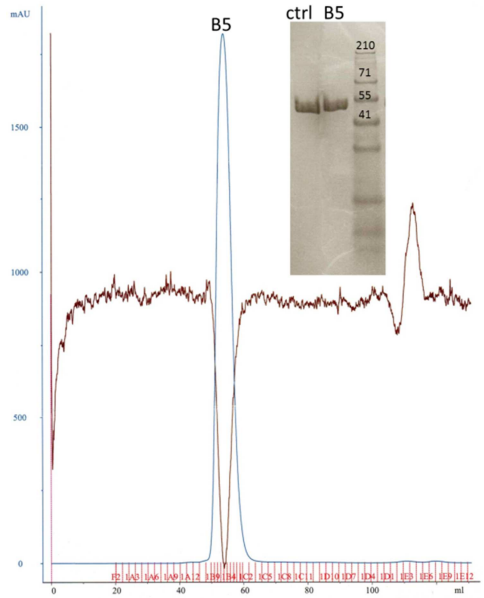


Figure 2.3 Chromatogram of *Gs*6PGDH on a Superdex 75 16/60 column after TEV – cleavage
The chromatogram displays the UV absorbance on the y-axis in blue in mAU, volume on the x-axis in ml, the conductivity of the solution in brown and in red the collection tubes. Fractions that were analyzed by SDS-page are labeled in the chromatogram accordingly. The SDS-page gel on the right shows the following from left to right: ctrl = defrosted, TEV - cleaved and purified 6PGDH enzyme as a control; sample from tube B5; protein ladder in kDa (SeeBlue)

Next, the pure protein was dialysed into crystallisation buffer C (50 mM Tris-HCl pH 7.4, 20 mM NaCl) and concentrated up to 12 mg/ml using VIVASPIN20 10,000 MWCO (from SartoriusStedim biotech). Aliquots were flash frozen in liquid nitrogen and stored at -20 °C.

Absorbance was measured at 280 nm to determine the enzyme concentration using a NanoVue Spectrophotometer. Enzyme concentration was calculated using the Beer-Lambert law:

$$A = \in lc$$

where $A = Absorbance,\ \in = molar\ extinction\ coefficient\ [L * mol^{-1} * cm^{-1}]$, $l = path\ length = 1cm\ and\ c = molar\ concentration$

The molar extinction coefficient was calculated based on the enzyme sequence (Table 2.3) using the ExPASy server (Table 2.4) (Artimo*, et al.,* 2012).

2.1.2 NMR fragment screen

The NMR fragment screen was carried out in collaboration with Daniel Fountaine (honours Student) who prepared the samples for screening. All spectra were acquired using a Bruker Avance 500 MHz Spectrometer with a 5 mm TXi cryoprobe at 298 K. For saturation transfer difference (STD) experiments (Meyer 1999; 1999) the sample was irradiated at 0.5 ppm. Water suppression was achieved by excitation sculpting (T.L. Hwang 1995). The wLOGSY experiment was carried out using ePHOGSY-NOE, a selective excitation with a 180° shaped pulse at the H_2O position or at another frequency based on a sequence written by Claudio Dalvit (Dalvit and Böhlen 1996; 1996; Dalvit*, et al.,* 2000; Dalvit 2009). The sample volume was a 500 µl aqueous suspension of 10 µM protein in phosphate buffer (10 mM Na_2PO_4, 1.76 mM KH_2PO_4) pH 7.5, 50 mM NaCl, 0.5 mM compound (dissolved in DMSO-d6) and 50 µl D_2O. In total 12 fragment compounds were mixed per tube and tested at once. As controls the reference spectra of the compounds were compared to the spectrum from the compound-mix in each tube. Subsequently, to test for

competition, the substrate 6PG was added in excess (1.25 mM) to the mixture and the spectra were recorded again.

2.1.2.1 NMR analysis

The spectra were analysed manually using TOPSPIN 2.1. First, the ^{1}H – spectrum was compared with each compound´s reference ^{1}H - spectrum to ensure every compound added was present. Then the 2D spectrum from the STD experiment was separated into two ^{1}H – spectra (in-house script) and analysed. Every signal found was compared to the ^{1}H reference spectrum. If a compound signal was present in the STD experiment, it was further analysed if it was shown as a binder in the wLOGSY spectrum. A compound was only considered a hit if:

- It had a signal in STD experiment
- It was shown to bind in the wLOGSY experiment
- It had a signal reduction in both experiments when 6PG was added

2.1.3 Kinetic characterisation of *Gs*6PGDH

All kinetic assays were carried out on a SPECTRA max 340PC (Molecular Devices). For all kinetic studies the His-Tag cleaved crystallisation construct (Table 2.3) was used.

For kinetic characterisation of *Gs*6PGDH a colorimetric assay was used, where the absorption of product NADPH at 340 nm was measured over time. The buffer was an aqueous mixture of 50 mM Tris – HCl at pH 7.4, 1 mM DTT, 250 mM NaCl and 0,02% CHAPS. The slope of the linear phase of each absorption curve was measured using a linear regression curve from Microsoft Excel where R^2 was > 0.9.

2.1.3.1 Assay development

To be able to accurately determine the inhibition of compounds the dose-response curve must be linear over the observed time frame. Because *Gs*6PGDH is from a thermophilic organism, all biochemical assays were performed at 30° C. At this temperature the reactivity of *Gs*6PGDH gave a sufficient signal to noise ratio (Figure 2.4) and the assay was easier to handle on the available instrument than at higher temperatures.

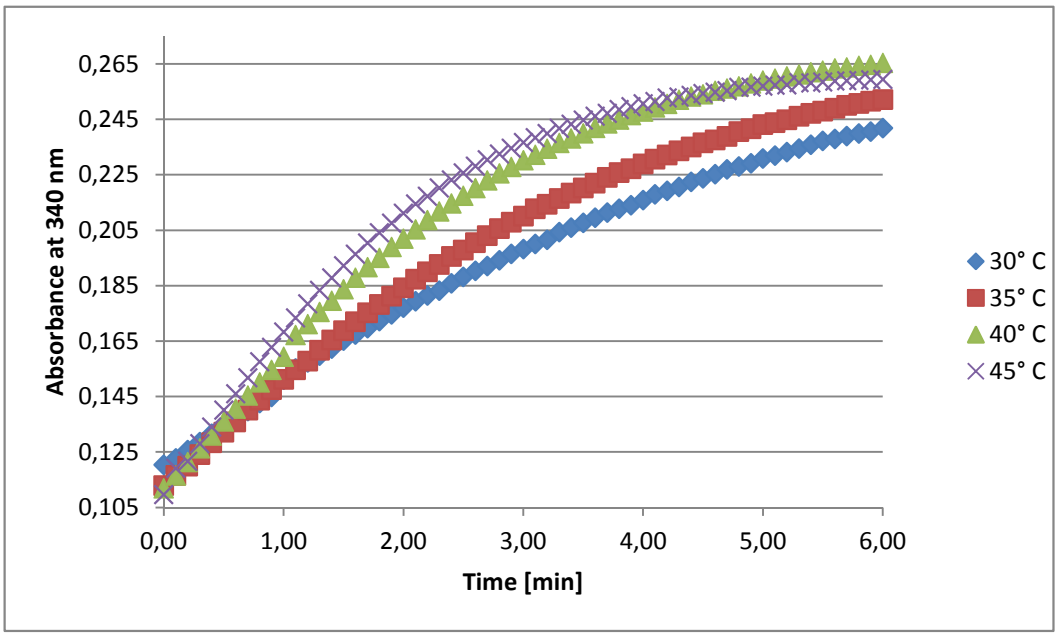


Figure 2.4 Temperature screen of *Gs*6PGDH
Time in minutes on x - axis versus absorbance at 340nm in atomic units (AU) on the y – axis is plotted at different temperatures. The blue curve was recorded at 30 ° C, red at 35° C, green at 40° C and purple at 45° C.

First, a $NADP^+$ standard curve was measured (Figure 2.5) to make sure that the measured signals linearly depends on the $NADP^+$ concentration.

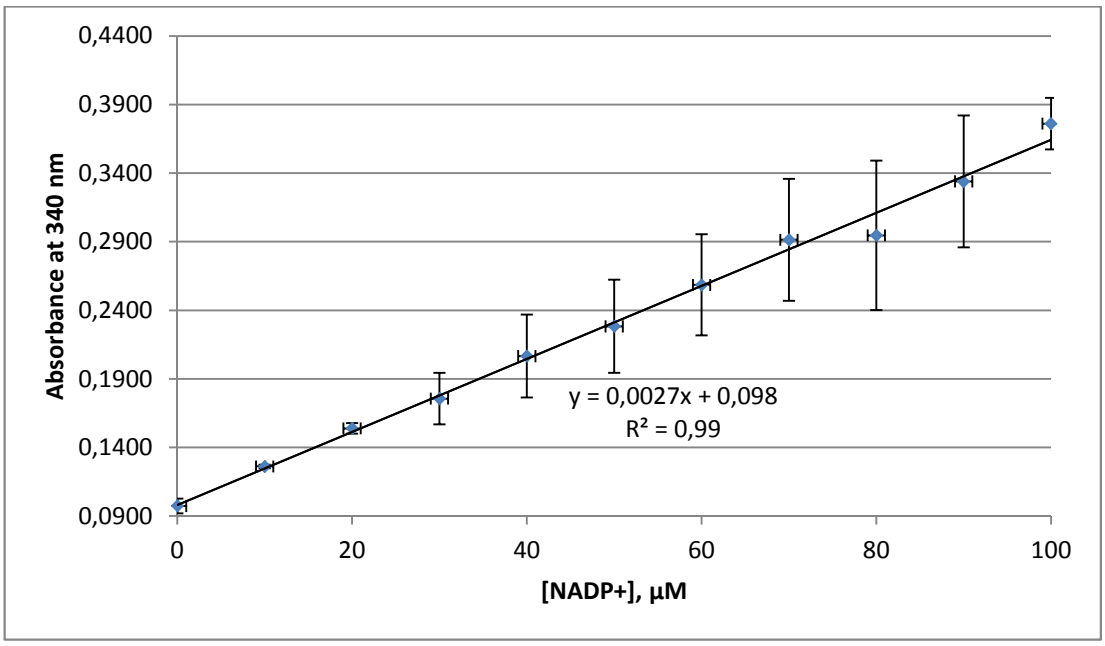


Figure 2.5 $NADP^+$ standard curve with a linear regression fit
$NADP^+$ concentration curve from 0 -100 µM on the x – axis versus absorbance at 340 nm on the y - axis and standard deviation as error bars (N = 2) with an R^2 for the linear regression of 0.99.

Further, the optimal *Gs*6PGDH concentration was determined by measuring the absorbance over time at different *Gs*6PGDH concentrations at fixed concentrations of 60 µM $NADP^+$ and 400 µM 6PG (Figure 2.6).

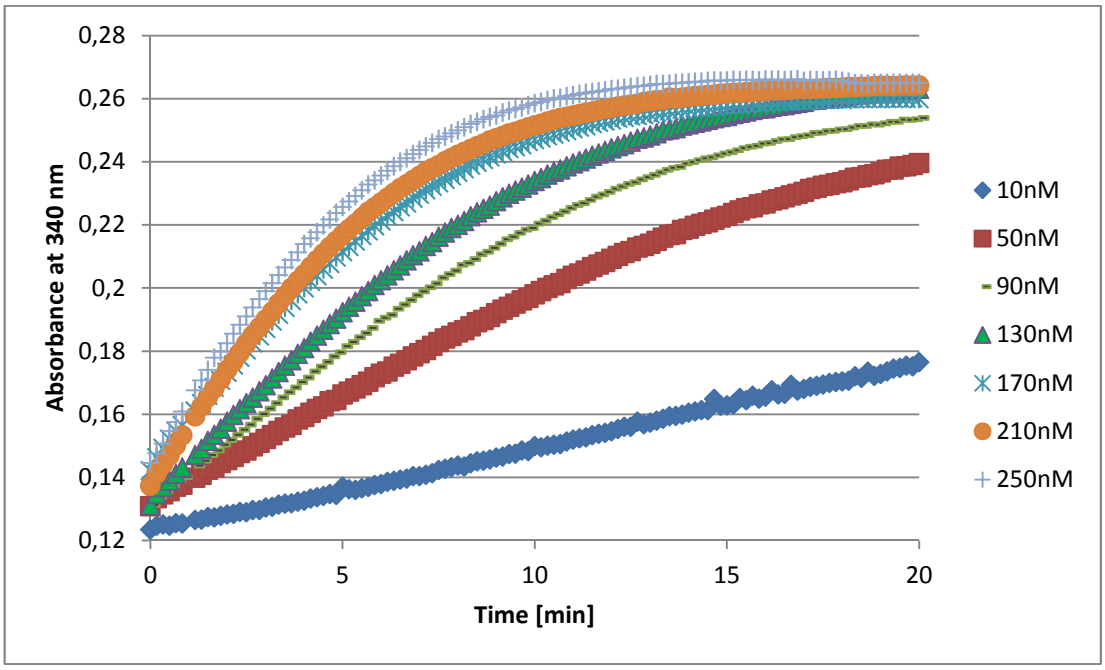


Figure 2.6 Reaction diagram at different *Gs*6PGDH concentrations
Reaction diagram at different enzyme concentration with time in minutes on the x – axis versus absorbance at 340 nm on the y - axis.

A *Gs*6PGDH concentration of 130 nM (Figure 2.6, green triangle) was found to result in a linear change in UV absorbance for 5 minutes and gave a sufficient signal to noise window for further measurements. Next, Michaelis - Menten kinetic was carried out to measure the affinity (K_m) of the substrate 6PG and the velocity (V_{max}) of the enzyme *Gs*6PGDH.

$$v = \frac{V_{max}[S]}{K_m + [S]}$$

Equation 2.1 Michaelis – Menten kinetic
Where v = reaction rate, [S] = substrate concentration, V_{max}= maximum rate of enzyme at saturated substrate concentration, K_m= Michaelis constant at which reaction rate is half maximum (substrate affinity).

$$\frac{1}{v} = \frac{K_m}{V_{max}[S]} + \frac{1}{V_{max}}$$

Equation 2.2 Lineweaver-Burk linearization of the Michaelis-Menten equation

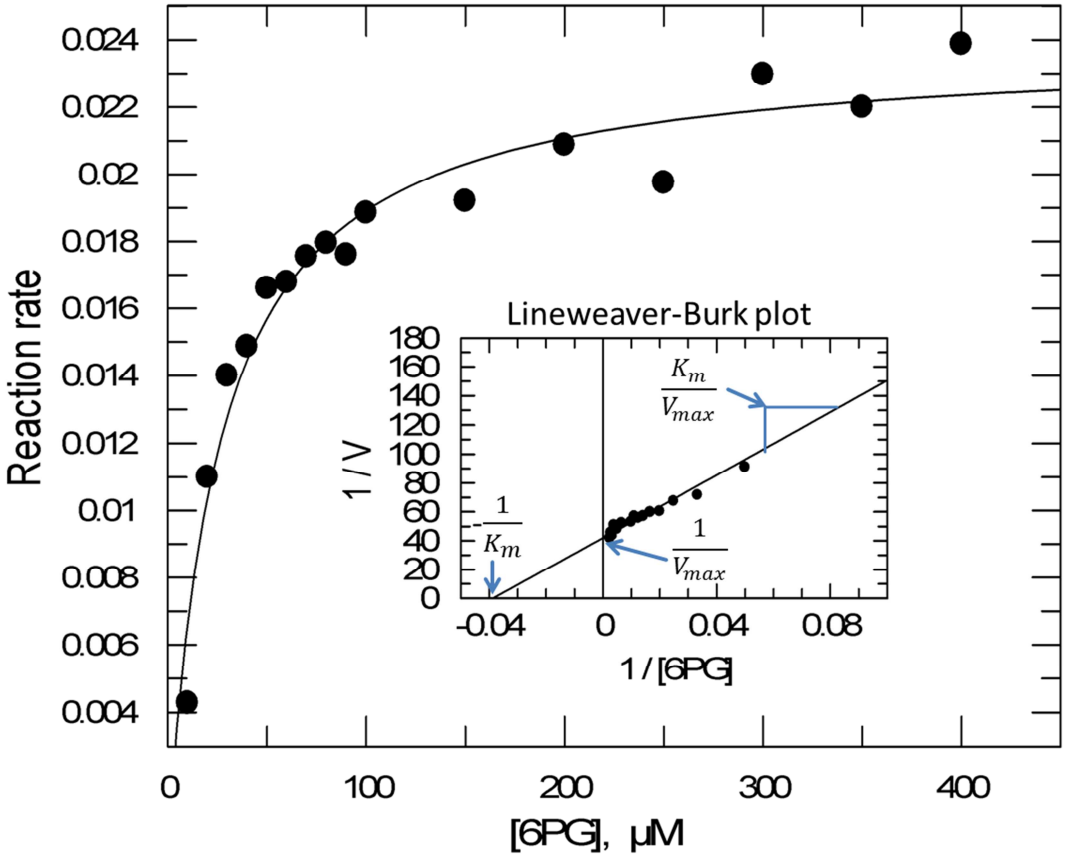


Figure 2.7 Saturation curve of *Gs*6PGDH with Lineweaver–Burk plot
Average saturation curve of *Gs*6PGDH from four measurements (N = 4) with substrate (6PG) concentration in µM on the x-axis and the reaction rate on the y–axis. As an insert a Lineweaver-Burk linearization (Equation 2.2) plot is shown. The x-intercept obtained by extrapolation of the positive experimental data is $-\frac{1}{K_m}$, $\frac{1}{V_{max}}$ is the y-intercept ($\frac{1}{[6PG]} = 0$) and $\frac{K_m}{V_{max}}$ is the slope of the line. The blue arrows indicate where

As a result a K_m of 25 ± 4 µM was determined for the substrate 6PG with an V_{max} of 0.024 ± 0.0009 µM/min. The value for K_m, V_{max} and their associated errors are the mean values from four independent measurements (N = 4).

For the inhibition assay the following concentrations were used: 130 nM *Gs*6PGDH, 25 µM 6PG (~ K_m) and 120 µM $NADP^+$ (in excess). The assays were carried out at 30 °C measuring absorption every 4 seconds until reaction was complete.

For the dose response curves two equations were used to fit the data points (both equations as implemented in GraFit from Erathacus Software Limited):

$$y = \frac{100\%}{1 + \left(\frac{x}{IC_{50}}\right)^s}$$

Equation 2.3 Two parameter fit equation
Where s = slope factor (gradient), x = cpd concentration, y = response.

$$y = \frac{Range}{1 + \left(\frac{x}{IC_{50}}\right)^s} + Background$$

Equation 2.4 Four parameter fit equation
Where s = slope factor (gradient), x = cpd concentration, y = response, *Range* is the fitted uninhibited value minus the *Background*. *Background* = the minimum y value multiplied by 0.95.

Equation 2.3 was used to fit data points which did not reach saturation at high inhibitor concentration or when not enough data points were available at high concentration. Otherwise, Equation 2.4 was used to fit the sigmoidal curve through the data points.

2.1.4 Crystallisation of *Gs*6PGDH

Crystals of Gs6PGDH were grown using the hanging drop method in 24-well plates on VDX slides (Hampton Research).

His-tag cleaved *Gs*6PGD (Table 2.3) was crystallised in two days in hanging drops consisting of 2.5 mM compound **1**.(compound number 12 from the following paper (Ruda*, et al.,* 2010)), 1 µl protein solution (12 mg/ml) and 1 µl reservoir (crystallisation buffer at pH = 7: 0.2 M sodium chloride, 2.5 mM $NADP^+$ and 2 M ammonium sulphate as a precipitant) equilibrated against 500 µl of reservoir at 20 °C as previously described (Cameron*, et al.,* 2009). The compound had to be dissolved in water at pH = 7 and added to the crystallisation buffer in order for the crystals to grow. Any DMSO impurities prevented crystal growth. Crystals were flash-cooled in crystallisation buffer with 20% glycerol.

2.1.4.1 Data collection, processing and structure modelling

X-ray diffraction data for the *Gs*6PGDH enzyme were collected at beamline ID14-1 in the European Synchrotron Radiation Facility (ESRF) in Grenoble (France), equipped with an ADSC Q210 CCD detector. A full data set of 720 images were collected from 321 ° - 141 ° with ϕ = 0.25 and 3 s exposure time. Crystal orientation, cell parameters and possible space group were determined using HKL2000 (Otwinowski and Minor 1997). Generated reflection lists and integrated reflections from the images were scaled and merged using SCALA (Evans 2006; 2011) from the CCP4 suite of programs (Potterton*, et al.,* 2003). Resolution data less than 38.8 or greater than 2.7 Å were excluded to produce an R_{merge} of 0.131.

Model generation was done using PHASER (McCoy 2007; McCoy*, et al.,* 2007). As a model enzyme the coordinates from *Gs*6PGDH (PDB 2W90) dimer were used as a starting point with a sequence identity set to 0.95 and a sequence file of the model enzyme 2W90. A solution was found with a rotation function score (RFZ) = 7.8 and a translation function score (TFZ) = 13.2. The structure was refined using REFMAC5 (Murshudov*, et al.,* 2011) from the CCP4 package. Refinement was using automatic weighting and local NCS restraints. Iterative model building was carried out using the interactive graphics program WinCOOT (v. 0.7.1 -pre) (Emsley and Cowtan 2004; Emsley*, et al.,* 2010). At first, all amino side chains with no electron density were mutated as stubs and after a refinement with REFMAC5 the amino side chains were mutated back and the occupancy of the atoms without electron density were set to 0.01. The structure was refined using the validation steps available in WinCoot and checked using web-based validation server MolProbity (http://molprobity.biochem.duke.edu/) (Chen*, et al.,* 2010) and RCSB PDB (http://validate.rcsb.org/).

2.2 Results

2.2.1 Structure analysis of 6PGDH enzymes

In the protein databank only one structure of *Tb*6PGDH (PDB 1PGJ) (Phillips*, et al.,* 1998) can be found (updated December 2013). This is due to its instability and therefore difficulties of overexpression and purification. For this reason the homologue enzyme *Gs*6PGDH was used for hit discovery due to its stability and binding site similarity. The binding sites of 6PGDH are highly conserved among species (Figure 2.8, Table 2.5).

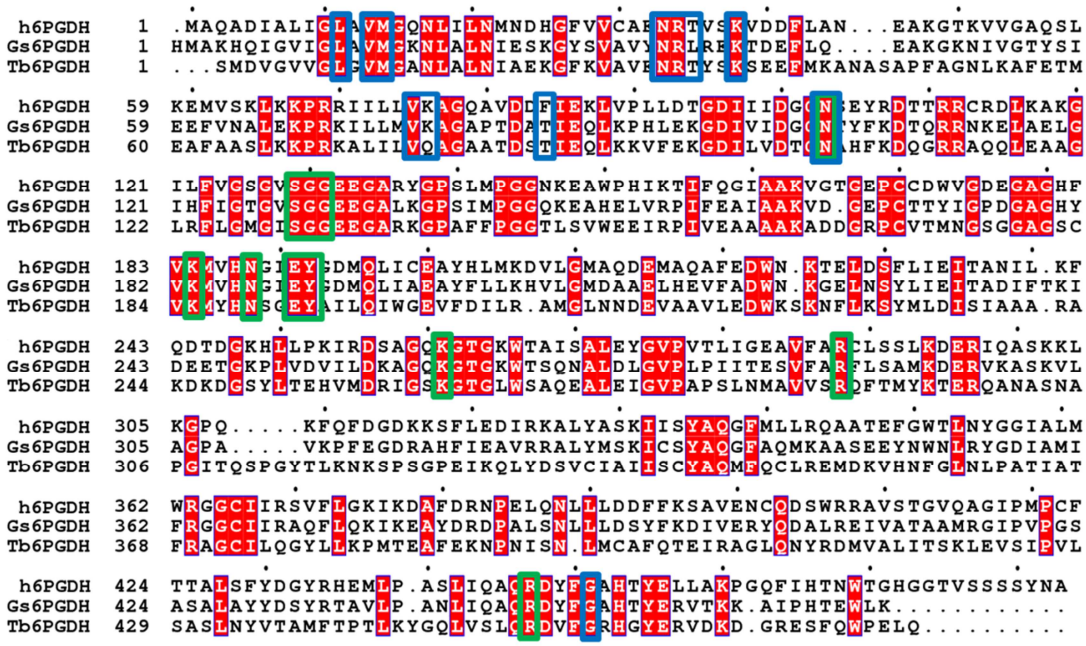


Figure 2.8 Sequence alignment of 6PGDH enzymes from different species
Sequence alignment from top: *Homo sapiens* (*h*6PGDH)*, Geobacillus stearothermophillus* (*Gs*6PGDH) *and Trypanosoma brucei* (*Tb*6PGDH). Identical amino acids are marked with a red background. Residues, which form a hydrogen bond to 6PG in the crystal structure 2W90 are marked with a green box and the ones that form a hydrogen bond to $NADP^+$ with a blue box.

Sequences were aligned using the web-based program ClustalW2 from the EMBL-EBI webpage (http://www.ebi.ac.uk/Tools/msa/clustalw2/)(Larkin*, et al.,* 2007). Similarity colouring was done using the web-based program ESPript 2.2 (http://espript.ibcp.fr)(Gouet*, et al.,* 2003).

Table 2.5 Sequence identity of 6PGDH enzymes from different species

Name	Length amino acids	Name	Length amino acids	Identity in %
*Gs*6PGDH	470	*Tb*6PGDH	478	36
*Gs*6PGDH	470	*h*6PGDH	483	56
*Tb*6PGDH	478	*h6PGDH*	483	32

Comparison of *Gs*6PGDH and *Tb*6PGDH sequences revealed residues conservation for the N-terminal, central and tail domains to 49 %, 26 %, 44 % respectively. The N-terminal and the tail domain contribute to the creation of the active site and show the highest level of conservation. The central domain contributes to dimer formation and is less well conserved. All structures of 6PGDH that are currently found in the protein database are dimers. The substrate 6PG does make defined hydrogen-bond interaction in the conserved binding site. The 6PG phosphate is accepting hydrogen bonds from Tyr190, Lys260 and Arg287. The hydrogen-bond interactions with Asn186, Thr262 (water mediated), Asn102, Ser128, Gly129 and Gly130 are holding the ligand in place. Two amino acids, which are important for the enzyme mechanism, Lys182 (Zhang*, et al.,* 1999) and Glu189 (Karsten*, et al.,* 1998), are also important for ligand binding (Figure 2.9).

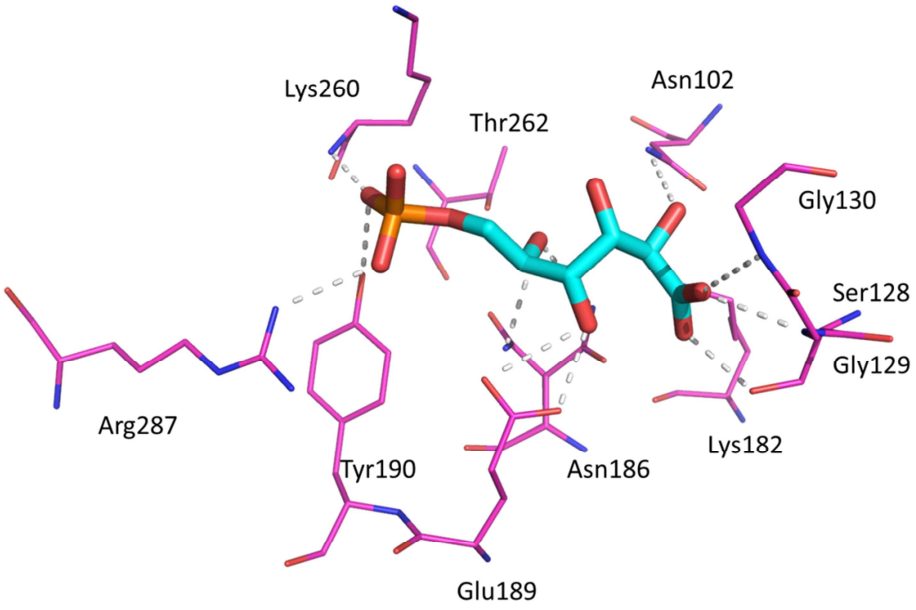


Figure 2.9 *Gs*6PGDH interactions with the substrate 6PG
In pink are the interacting *Gs*6PGDH amino acids making hydrogen bonds (white dots) to 6PG (lightblue carbon atoms). For clarity water mediated interactions are not shown.

For a better comparison of the $NADP^+$ binding sites of *Gs*6PGDH, *Tb*6PGDH and *h*6PGDH, homology models of the enzymes were created using the online pipeline SWISS MODEL (http://swissmodel.expasy.org/)(Arnold*, et al.,* 2006). As a result of the algorithm in the underlying homology building program, whenever possible, the backbone of the build homology models traces exactly that of the template and positions of conserved amino acid side chains are not optimized in the models. For the purpose of analysing the $NADP^+$ binding sites this has the advantage that differences in the amino acid residues can be easily spotted by visual examination (Figure 2.11).

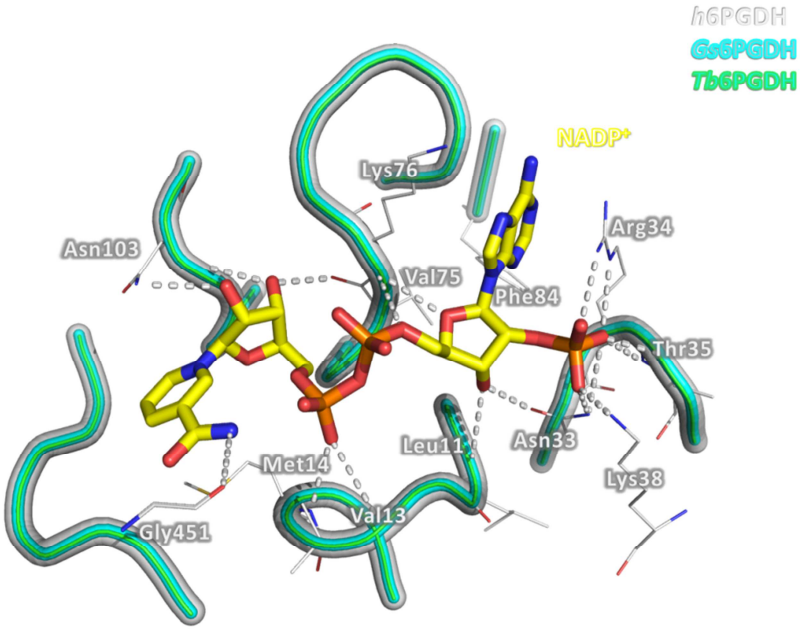


Figure 2.10 Overlay of the NADP$^+$ binding pocket from *h-, Gs- and Tb*6PGDH
Carbon atoms of *h*6PGDH are marked in grey, of *Gs*6PGDH in light blue and of *Tb*6PGDH in green. The carbon atoms of cofactor NADP$^+$ is showed in yellow with hydrogen bonds (shown as white dots) to *h*6PGDH residues (shown as grey text). For clarity the overlay shows only part of the pocked at 7 Å around NADP$^+$ and only side chains of *h*6PGDH are shown.

The overlay of *h*6PGDH and the homology models of *Gs*6PGDH and *Tb*6PGDH show a perfect conservation of the backbone chain but reveal several differences in the side chains of the enzymes. All the different amino acids around NADP$^+$ are shown in Figure 2.11 and listed in detail in Table 2.6. The NADP$^+$ phosphate is accepting hydrogen bonds from Thr35, Lys38 and Arg34. The hydrogen-bond interactions with Phe84, Leu11, Val75, Met14, Val13, Asn103 and Gly451 are holding the co-ligand in place.

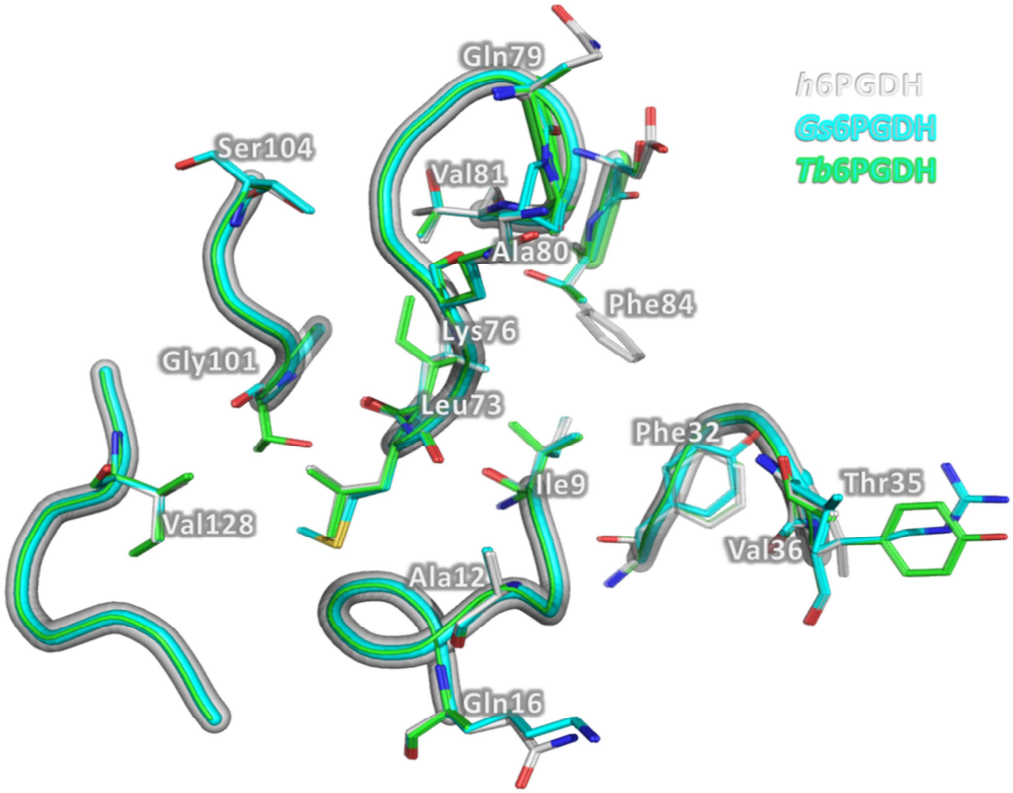


Figure 2.11 Overlay of the NADP$^+$ binding pocket from *h-, Gs-* and *Tb*6PGDH showing differences of side chains.
Carbon atoms of *h*6PGDH are marked in grey, of *Gs*6PGDH in light blue and of *Tb*6PGDH in green. For clarity the overlay shows the same orientation as Figure 2.10 but without NADP$^+$ and only the different side chains of *h*6PGDH are shown. For the corresponding amino acids of *Gs*- and *Tb*6PGDH see Table 2.6

All amino acids with side chains oriented to NADP$^+$ and which are different among the species can cause selectivity. For example the positively charged Lys76 in *h*6PGDH is Gln77 in *Tb*6PGDH which means, that a positive charged group is replaced by an uncharged side chain. The polar side chain of Gln79 in *h*6PGDH is a hydrophobic Ala80 of *Tb*6PGDH. Especially interesting is the difference of Phe84 in *h*6PGDH to Thr85 in *Tb*6PGDH. Here not only a hydrophobic side chain is changed to

a polar one, but also the aromatic ring system of Phe84 can do stacked interaction with the selective inhibitor.

Table 2.6 Differences in side-chain - amino acids of *h*-, *Tb*- and *Gs*6PGDH at 7 Å around $NADP^+$

*h*6PGDH	*Gs*6PGDH	*Tb*6PGDH
Ile9	Ile11	Val7
Ala12	Ala14	Gly10
Gln16	Lys18	Ala14
Phe32	Tyr33	Phe30
Thr35	Leu36	Thr33
Val36	Arg37	Tyr34
Leu73	Leu73	Ile74
Lys76	Lys76	Gln77
Gln79	Ala79	Ala80
Ala80	Pro80	Ala81
Val81	Thr81	Thr82
Asp83	Ala83	Ser84
Phe84	Thr84	Thr85
Gly101	Gly101	Thr102
Ser104	Thr104	Ala105
Val128	Val128	Ile129

Considering the fact, that the hydrogen-bond pattern of 6PG is conserved among species and that the cofactor site show several differences, the strategy was:

1. To find compounds that bind into the 6PG binding site using *Gs*6PGDH as the thermo stable Model-enzyme
2. To chemically modify the compound found by the first stage so that it will expand into the $NADP^+$ pocket and bind selectively to *Tb*6PGDH exploiting the differences in the $NADP^+$ binding site.

It was previously shown that homologue enzymes can be used to study binding modes of *Tb*6PGDH inhibitors (Sundaramoorthy*, et al.,* 2007). Sundaramoorthy *et al.,* (2007) used *Ll*6PGDH to determine the binding mode of a HEI inhibitor (**4)** by using crystallisation methods.

The substrate binding site conservation between *Gs*6PGDH and *Tb*6PGDH and its increased stability made *Gs*6PGDH a good model enzyme for hit discovery. Additionally, it was further shown, that *Gs*6PGDH can be crystallized and therefore used for X-ray binding mode determination of potential inhibitors.

2.2.2 Inhibition assay with virtual screening hits

A virtual screening was carried out previously (Ruda*, et al.,* 2010) and from this, 18 *Tb*6PGDH inhibitors were identified. The binding modes of these inhibitors were not determined, which would be a great help to establish a SAR and produce lead compounds. Therefore, the homologue enzyme *Gs*6PGDH was used to determine a

binding mode for the virtual screening hits. Additionally, the virtual screening hits should be kinetically tested for activity. Therefore, all virtual screening hits were tested against *Gs*6PGDH and the active compounds were planned to be tested against *Tb*6PGDH. As a control compound a high energy intermediate hydroxamate **4** (Dardonville*, et al.,* 2004) was chosen due of its low K_i of 0.01 µM for *Tb*6PGDH.

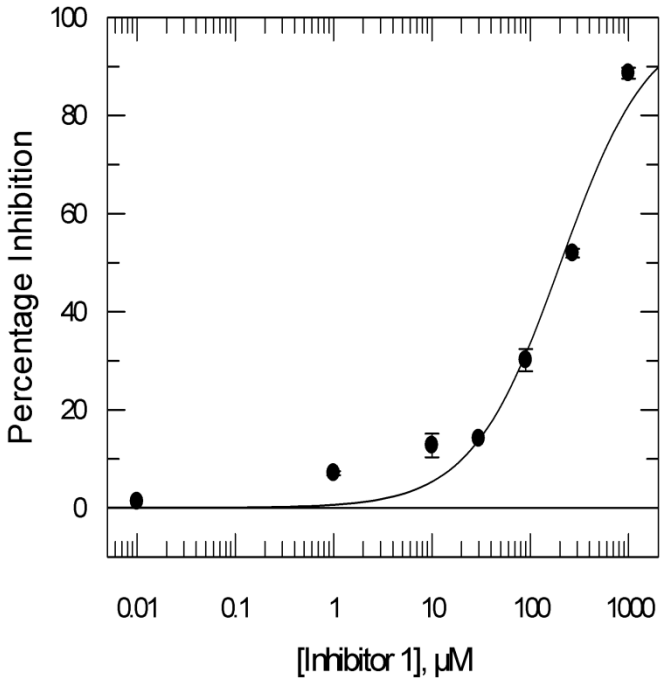


Figure 2.12 Dose response curve of 1
Dose-response curve with error bars for each data point from two experiments (each with n=2), percentage inhibition on y-axis and compound concentration in µM on x-axis (logarithmic scale). Points were fitted using a two parameter equation (Equation 2.3). Analysis from two experiments returns a pIC_{50} = 3.69±0.02 and a Hill slope of 1.0 ± 0.1 (mean ± SD; N=2).

Only three compounds out of 18 were found to inhibit *Gs*6PGDH (Table 2.7; compound **1** see Figure 2.12, compound **2** see Figure 2.13 and compound **3** see Figure 2.14). Except for compound **2**, the pIC_{50} values were lower as the published ones for *Tb*6PGDH (Ruda*, et al.,* 2010). For **1** a pIC_{50} of 3.69±0.02 (published 4.29 for

*Tb*6PGDH), for **2** a pIC_{50} of 4.37±0.05 (published 4.28 for *Tb*6PGDH) and for **3** a pIC_{50} of 3.43±0.04 (published 4.36 for *Tb*6PGDH) (Figure 2.15). The largest discrepancy was found for the control compound **4**. The published K_i of **4** for *Tb*6PGDH was 0.01 µM, but the calculated K_i based on the measured IC_{50} value (Equation 3.3) for *Gs*6PGDH was 4.2±1.2 µM which is about four hundred times higher.

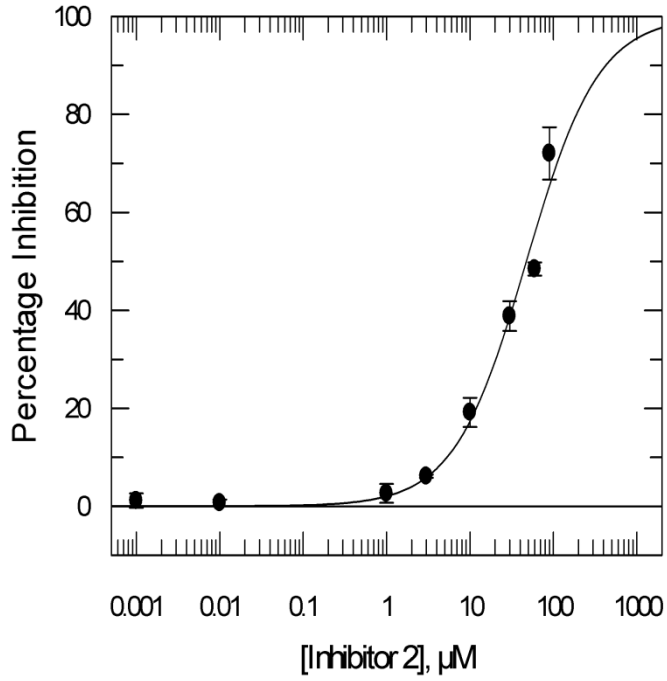


Figure 2.13 Dose response curve of 2
Dose-response curve with error bars for each data point from three experiments (each with n=2), percentage inhibition on y-axis and compound concentration in µM on x-axis (logarithmic scale). Points were fitted using a two parameter equation (Equation 2.3). Analysis from three experiments returns a pIC_{50} = 4.37±0.05 and a Hill slope of 1.0 ± 0.15 (mean ± SD; N=3).

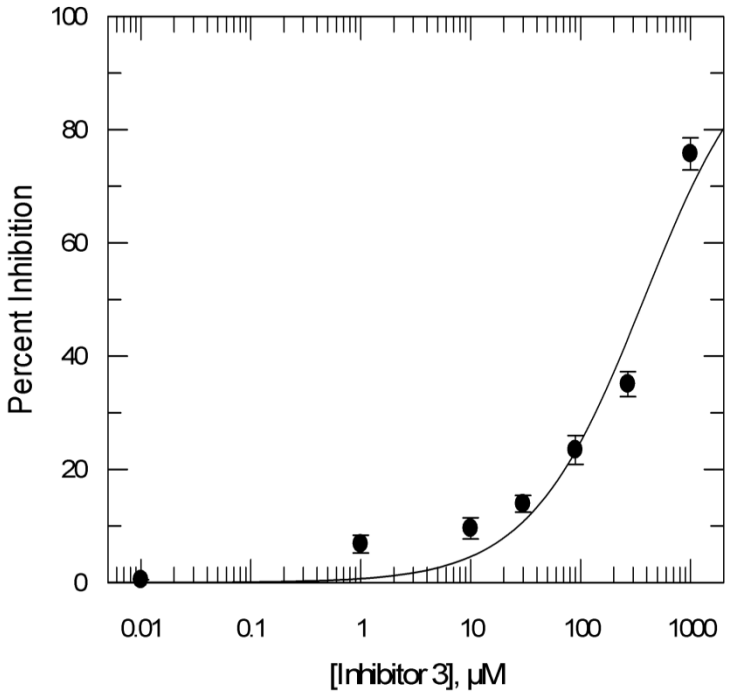


Figure 2.14 Dose response curve of 3

Dose-response curve with error bars for each data point from two experiments (each with n=2), percentage inhibition on y-axis and compound concentration in µM on x-axis (logarithmic scale). Points were fitted using a two parameter equation (Equation 2.3). Analysis from two experiments returns a pIC_{50} = 3.43±0.04 and a Hill slope of 0.8 ± 0.00 (mean ± SD; N=2).

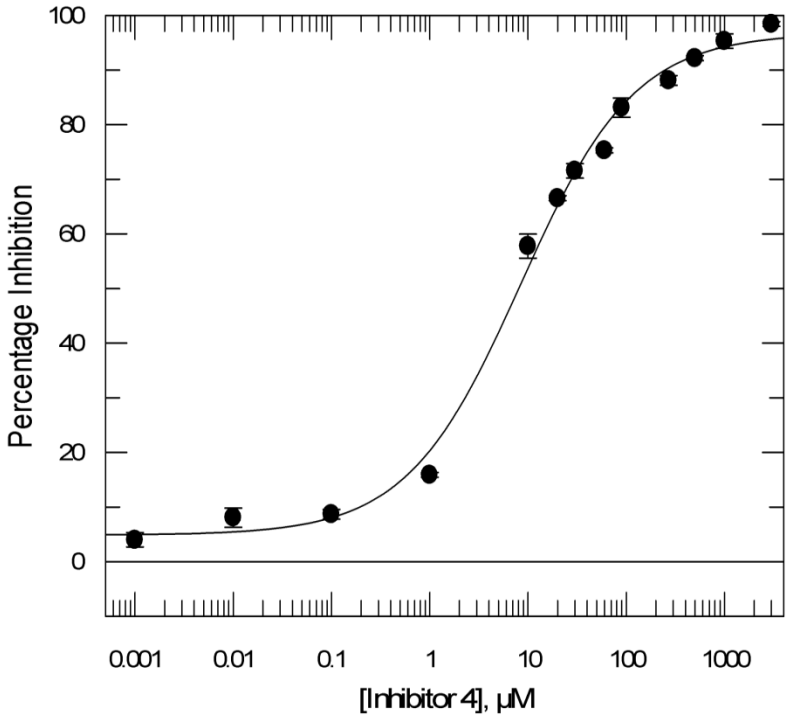


Figure 2.15 Dose response curve of 4

Dose-response curve with error bars for each data point from two experiments (each with n=2), percentage inhibition on y-axis and compound concentration in µM on x-axis (logarithmic scale). Points were fitted using a four parameter equation (Equation 2.4). Analysis from two experiments returns a pIC_{50} = 5.07±0.02 and a Hill slope of 0.8 ± 0.02 (mean ± SD; N=2).

Table 2.7 Inhibition data for virtual screening hits published in (Ruda*, et al.,* 2010)

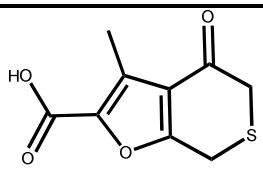
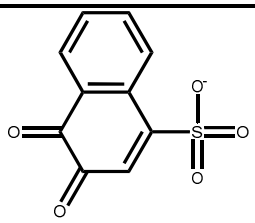
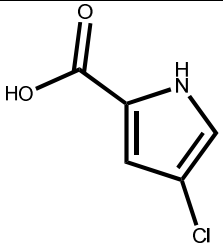
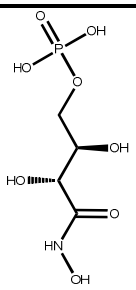
#	Structure	published *Tb*6PGH IC_{50} converted to pIC_{50}	published *Tb*6PGH Hill slope	*Gs*6PGDH pIC_{50}	*Gs*6PGDH Hill slope
1		4.29	1.7	3.69±0.02	1.0±0.1
2		4.28	4.8	4.37±0.05	1.0 ± 0.15
3		4.36	1.8	3.43±0.04	0.8 ± 0.00

Table 2.8 Inhibition data for high energy intermediate analogue as described by (Dardonville*, et al.,* 2004)

#	Structure	published *Tb*6PGH K_i [µM]	published *Tb*6PGH Hill slope	*Gs*6PGDH[a] K_i [µM]	*Gs*6PGDH Hill slope
4		0.01	N/A	4.2±1.2	0.8±0.02

[a] **K_i was based on the determinded IC_{50} value (Equation 3.3) for a better comparison to the published data.**

2.2.3 Determination of binding mode of virtual screening hits using crystallography

The crystal structure of *Gs*6PGDH was crystallised with the inhibitors as described previously (Cameron*, et al.,* 2009). The crystals only grew if compound **1** and $NADP^+$ were present. Diffraction was calculated in-house on a Rigaku X-ray instrument with R-AXIS IV++ imaging plate with a resolution of 2.7 Å. Data for same crystal were collected at ESRF in Grenoble with a resolution of 1.8 Å (Table 6.3). After refinement, no electron density for the ligand was found.

Virtual screening compound **1** was the only one that was suitable to further optimisation. Every attempt to crystallise *Gs*6PGDH without a ligand or just with

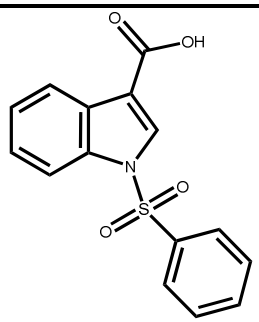
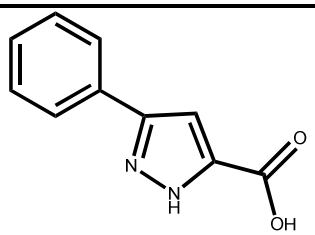
$NADP^+$ was unsuccessful, so therefore no apo-crystals could be produced for ligand soaking. When compound **1** was dissolved in DMSO, no crystals could be grown under the same conditions as published by Cameron, S. *et al.,* 2009. Only by adjusting the pH to 7, compound **1** could be solubilised in crystallisation buffer without DMSO. This adjustment allowed *Gs*6PGDH to crystallise in presence of 2.5 mM **1** and 2.5 mM $NADP^+$.

2.2.4 Fragment screening using NMR methods

The in-house fragment library containing 652 fragment-like compounds was screened for binding to *Gs*6PGDH as described (2.1.1). From this library 36 compounds showed binding in both NMR experiments (STD and wLOGSY) and showed a signal reduction after adding 1.25 mM 6PG for competition. None of the 36 compounds had a carboxylate group, which is surprisingly considering that all virtual screening hits found by Ruda, G.F. *et al.,* (2010) were acids. Therefore, two acids (**5** and **6**) from the screen were also short-listed for further studies (Table 2.9). Acid **5** was a binder in STD and wLOGSY experiment, but a signal reduction after adding 6PG could not be seen in STD. The second acid **6** showed only binding in STD without a signal reduction with 6PG. Despite these hits, 100 other compounds showed binding in at least one of the experiments. That means that 20.9% (136 out of 652) of all screened compounds bound to *Gs*6PGDH, indicating a high false positive rate. This high rate can be explained in parts by the weak binding of 6PG to

*Gs*6PGDH (K_m =25 µM) which makes it problematic to achieve competition with a potential inhibitor. An inhibitor in nM range would have been a more reliable indicator for the hit identification, but was unfortunately not available.

Table 2.9 Active acids found by a fragment screening for *Gs*6PGDH

#	Structure
5	
6	

2.2.5 Inhibition assay with NMR fragment hits

All 38 NMR hits were tested for *Gs*6PGDH inhibition at 1 mM compound concentration in a biochemical assay which was previously described (2.1.3). Due to no or only weak inhibition at this concentration, it was not possible to measure a dose response curve for either of the compounds. Only **5** (PI = 50% at 1 mM) and **6** (PI = 32% at 1 mM) were active at the chosen concentration (Table 2.9). All of them

were acids. The reason for this is unclear. However, this result is consistent with the published data (Ruda*, et al.,* 2010). All compounds found by Ruda*, et al.,* 2010 in a virtual screening were acids, too. Crystallisation trials with acids found by NMR fragment screening

Two active acids (**5** and **6**) were set up for crystallization trials as previously described (2.1.4). Unfortunately none of the acids crystallised under the previous conditions.

2.3 Discussion

2.3.1 Virtual screening hits

Most virtual screening hits found by Ruda, G.F. *et al.,* 2010 were not active against *Gs*6PGDH (Table 2.7). Unfortunately the publication does not mention important details on how the inhibition screen was carried out. This fact makes it difficult to evaluate *Gs*6PGDH as a model enzyme for *Tb*6PGDH. The following important questions should have been addressed:

- At which temperature was the screen carried out? This is especially important to know, because of the instability of the *Tb*6PGDH enzyme as mentioned before (Phillips*, et al.,* 1998; Sundaramoorthy*, et al.,* 2007).
- How many replicates were carried out (N = ?) and how was the error calculated?
- What model was used to fit the parameters and how were they normalised?
- Why the Hill slopes of all compounds, except for compound 13 (paper), were higher than 1?
- Some of the compounds do absorb at 340 nm or interact with the assay. How was inhibition tested for such compounds?
- What was used as a positive control for the inhibition?

Here, for inhibition studies of *Gs*6PGDH the high energy intermediate **4** was used as a positive control. The published potency for this compound (Dardonville*, et al.,*

2004) could not have been confirmed. The paper does as well not mention important points:

- How was the stability of *Tb*6PGDH at 20° C ensured while the inhibition assay was carried out?
- How many replicates were carried out (N = ?) and what is the error of K_i and IC_{50}?

As a conclusion two scenarios could be possible.

1. The conserved 6PG binding site is not enough to translate the binding affinity of the compounds from *Tb*6PGDH to *Gs*6PGDH. It is also plausible that the compounds bind into the less conserved NADPH binding site instead of the 6PG site. Consequently there must be other structural factors than the hydrogen-bond interaction pattern, which causes the differences in binding affinities. While difficult to rationalise, such a behaviour has already been observed for other enzymes (Baba*, et al.,* 2003; Cleghorn*, et al.,* 2011; Teng*, et al.,* 2013). Nevertheless, compounds **1, 2, 3** and **4** were confirmed as inhibitors, even though not as active as for *Tb*6PGDH.
2. The published data on which this study was based on is not correct and needs to be reanalysed considering the questions addressed above.

2.3.2 Crystallisation and crystal structure determination

The crystals diffracted at 2.7 Å in-house and at 1.8 Å at the synchrotron. The analysis of the X-ray structure showed no ligand in the active site. All previous crystallisation experiments have shown that crystals only appear in the present of the ligand or the substrate. One reason, that the ligand was not present in the crystal structure could be the transfer of the crystal from the mother liquor to 20% glycerol for cryo protection. Every direct transfer of the crystal to the cryoprotectant caused the crystal to dissolve. Therefore, a step-wise transfer to 10%, 15% and 20% glycerol was necessary. Even though the ligand was present in each reservoir during the transfer, the low binding affinity may have caused the ligand to diffuse away from the crystal, especially as the cry-buffers are more hydrophobic than the buffer used for determining the affinity of the compounds. This reduces the contribution of the hydrophobic effect to binding affinity and is likely to result in weaker binding affinity in the cryo-conditions.

2.3.3 Fragment NMR – screening and hit validation

Our in-house compound library consisting of 652 diverse compounds was screened for binding affinity to *Gs*6PGDH using NMR. This approach identified 36 compounds able to bind the target protein in both experiments (STD and wLOGSY) with a decreasing signal after the addition of 6PG. From 38 selected compounds, only two acids were active at 1 mM in biochemical colorimetric assays which were selected

on a rational basis. One reason why 36 compounds were not confirmed by this method could be that they are all weak binders with an IC_{50} over 1 mM.

Both actives compounds were acids (**5** -> PI 50% and **6**-> PI 32%). At concentrations ≥ 500 µM the absorbance of the reaction solution with compound **5** increased, suggesting that either the compound absorbs at 340 nm or does react with $NADP^+$ producing more NADPH.

In summary, despite of all efforts, no binding mode could be determined of the virtual screening hits.

The high polarity of the active site and the lack of confirmed, competitive inhibitors questions 6PGDH as a druggable enzyme, at least for a rational hit discovery approach.

3 Hit discovery for UGP

3.1 Material and Methods

All reagents, unless otherwise stated were purchased from Sigma Aldrich.

Screening compounds **8, 9, 10** and **15** were purchased from Sigma Aldrich, **11** and 13 from Apollo Scientific and **20**, **21**, **22** from (Otava, Ltd).

3.1.1 Homology model

Sequence alignments between *Tb*UPG and *h*UGP were generated using ClustalW (Thompson*, et al.,* 1994). Subsequently, Modeller 9.2 (Sali and Blundell 1993) was used to build homology models of *h*UGP, whereas the *Tb*UGP crystal structure (PDB code 3GUE) served as a template. Modeller was run with default settings, and only the highest-scoring structure was used for further analysis and modelling.

3.1.2 Virtual screening and molecular docking

3.1.2.1 Compound database

From an in-house database of 5.2 million commercial available compounds with precalculated physiochemical properties (Brenk*, et al.,* 2008), only compounds which had the following properties were selected for virtual screening:

- Number of heavy atoms < 24
- Rotational bonds ≤ 4
- Number of hydrogen-bond acceptors ≥ 2
- Number of hydrogen-bond donors ≥ 1

3.1.2.2 Pharmacophore filter

The compounds selected for virtual screening were further filtered by a 3D pharmacophore defined using Unity (Tripos Inc., St. Louis, MO). The compounds were first converted to a 3D UNITY database using default parameters and macro files. The 3D pharmacophore was derived based on the interactions of the uracil moiety of the substrate as found in the crystal structure *Tb*UGP-substrate complex (PDB code 3GUE, Figure 3.1). Spheres with a 1.5 Å radiuses were placed around the positions of the nitrogen atom of the uracil moiety and the carbonyl groups to indicate locations of hydrogen donor and acceptor groups, respectively. To consider the directionality of the hydrogen bonds the groups were connected with spheres placed around the amino group of Gly189 (donor), the carbonyl group of Gln161 (acceptor) and the amino group Gly83 (donor). Additionally to the hydrogen bond interactions a sphere indicating an aromatic group was placed at the centre of the uracil ring.

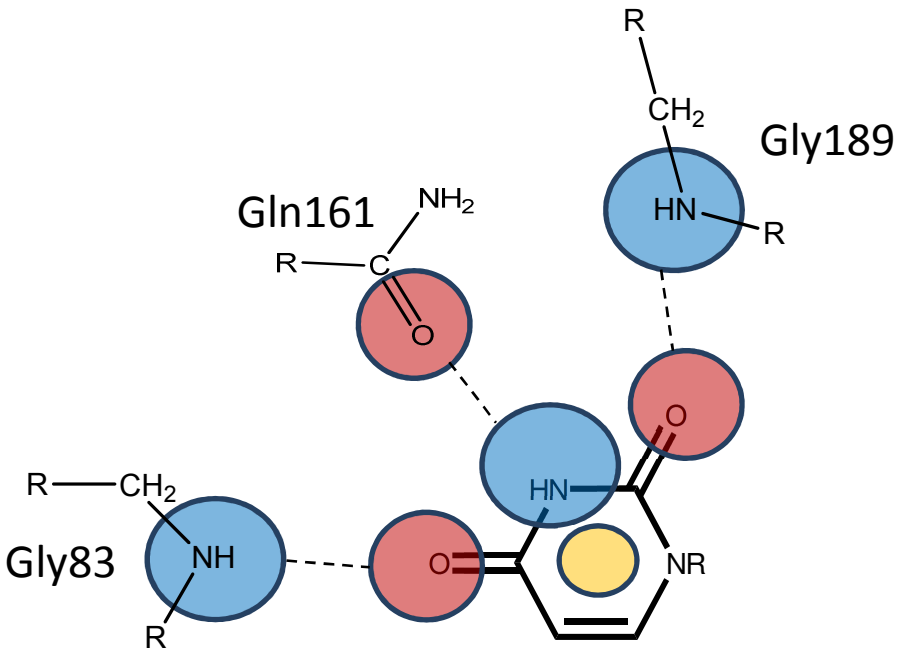


Figure 3.1 Schematic view of the 3D pharmacophore for potential *Tb*UGP inhibitors
Acceptor groups are deciphered as red cycles, donor groups as blue cycle, the aromatic feature is marked as a yellow circle and dotted lines indicate interactions between acceptor and donor groups.

The hydrophobic feature and at least two out of the possible three hydrogen bond interactions were required to fulfil the pharmacophore.

3.1.2.3 Receptor preparation for docking

For docking, the software DOCK 3.5.54 (Lorber and Shoichet 1998; Wei*, et al.,* 2002; Mysinger and Shoichet 2010) was used. The *Tb*UGP structure with the coordinates (PDB code 3GUE) (Marino*, et al.,* 2011) was used as receptor for docking. Hydrogen atoms were added using MOLOC (Gerber, molecular design) and their positions minimized using the MAB force field (Gerber and Muller 1995) as implemented in

MOLOC in presence of the product present in the structure, while keeping all non-hydrogen atoms rigid. Subsequently, all ligands, water molecules and ions were removed from the crystal structure. The sphere set used to define the region of the binding site with a low dielectric constant, was created based on bound UDP-Glc (3GUE). The product UDP-Glc was modified by adding atoms to the 2-hydroxy group of ribose to completely fill the cavity. For the sphere set, used for placing the ligands into the binding site, the matching atoms were placed around the three functional groups of uracil groups that are part of the pharmacophore (Figure 3.1 and Figure 3.2). Partial charges for all receptor atoms were obtained using the AMBER-99 force field parameters (Wang 2000). The electrostatic potential was calculated using DelPhi (Nicholls and Honig 1991), with a grid size of 65 and an internal dielectric constant of 2 and an external dielectric constant of 78. The van der Waals potential was calculated using a DOCK utility named CHEMGRID (Meng*, et al.,* 1992; Shoichet*, et al.,* 1992). Maps to calculate partial ligand desolvation were generated using SOLVMAP (Mysinger and Shoichet 2010).

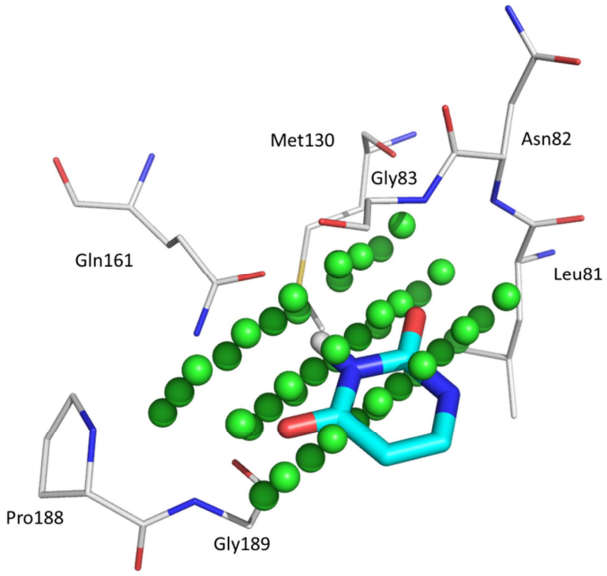


Figure 3.2 Matching sphere set generated for UGP
The green spheres indicate the positions of the matching points used for docking. The uracil ring is shown for clarity, where carbon atoms are indicated in light blue, nitrogen atoms in dark blue, oxygen atoms in red and part of the surrounding enzyme with white carbon atoms.

3.1.2.4 Small molecule preparation and molecular docking

DOCK 3.5.54 was used to dock small molecules flexibly in to the active site of *Tb*UGP. The small compounds with pre-calculated physicochemical properties selected from the database (3.1.2.1), were filtered as described (3.1.2.2). All compounds were stored in SMILES format (Weininger 1988). The protonation- and tautomeric states for each compound were calculated using an in-house script based on the OEToolkit (Openeye, Santa Fe, NM). The 3D and multiple low energy conformations were generated using OMEGA2 (Openeye, Santa Fe, NM). AMSOL (http://comp.chem.umn.edu/amsol/; (Wei*, et al.,* 2002)) was used to calculate the

desolvation energies and partial charges. Finally all ligands were aligned on their ring systems and stored in a hierarchical format (Wei*, et al.,* 2002).

The following settings were chosen to sample ligand orientations: ligand and receptor bins were set to 0.5 Å, and overlap bins were set to 0.4 Å; and the distance tolerance for matching ligand atoms to receptor matching sites ranged from 1.1 to 1.2 Å. Only docking poses which did not placed any atoms in areas occupied by the receptor was scored for electrostatic interaction energy and van der Waals and complementarity (Lorber and Shoichet 1998) and penalized according to its estimated partial desolvation energy. For each compound, only the best-scoring database representation (tautomer, protonation state, multiple ring alignment) was stored in the final docking hit list.

3.1.2.5 Docking analysis

All docked compounds were visually inspected using Pymol (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC). Only compounds which were placed in the uracil binding site and made at least two hydrogen bonds to the protein were selected for further investigations. If the compounds contained moieties exceeding the uracil binding site, it was required that they pointed into the cavity located opposite the 2-hydroxy group of the ribose in the *Tb*UGP complex structure (Figure 3.8 a)). This was mandatory in order to search for selective compounds that would not inhibit *h*UGP.

3.1.3 Protein overexpression and purification of *Tb*UGP

Two different constructs A and B of *Tb*UGP were used: For kinetic assays the "kinetic assay construct" A (Table 3.1) and for crystallisation the "crystallisation construct" B (Table 3.3). Both constructs were purified as described before (Marino*, et al.,* 2010).

Table 3.1 Amino acid sequence of *Tb*UGP kinetic assay construct A

MGSSHHHHHHSSGLEVLFQGPHMPLNPPSAFSGAALACLEKMQASGVEEKCIHIFLIQHALVR
KGETGYIPEKSISPVESLPFLQGIETKGENTALLRQAVVLKLNGGLGTGMGLNGPKSLLQVKNG
QTFLDFTALQLEHFRQVRNCNVPFMLMNSFSTSGETKNFLRKYPTLYEVFDSDIELMQNRVPKI
RQDNFFPVTYEADPTCEWVPPGHGDVYTVLYSSGKLDYLLGKGYRYMFISNGDNLGATLDVRL
LDYMHEKQLGFLMEVCRRTESDKKGGHLAYKDVIDETTGQTRRRFVLRESAQCPKEDEDSFQ
NIAKHCFFNTNNIWINLMELKKMMDEQLGVLRLPVMRNPKTVNPQDSQSTKVYQLEVAMG
AAISLFDRSEAVVVPRERFAPVKTCSDLLALRSDAYQVTEDQRLVLCEERNGKPPAIDLDGEHYK
MIDGFEKLVKGGVPSLRQCTSLTVRGLVEFGADVSVRGNVVIKNLKEEPLIIGSGRVLDNEVVV
VE

Table 3.2 Biochemical properties of *Tb*UGP kinetic assay construct A

Number of amino acids: 507

Molecular weight: 56857.1 Dalton

Theoretical pI: 6.23

Abs 0.1% (=1 g/l): 0.560 at 280 nm (Artimo*, et al.,* 2012)

Procedure

Cultures of *E. coli* BL21(DE3) cells containing a modified pET15B plasmid were grown in an incubator (INFORS HT Multitron) at 37 °C in LB medium supplemented with ampicillin (50 µg/ml) until OD_{600} of 0.9 was reached. The temperature was then lowered to 16 °C and the protein expression was induced with 0.5 mM IPTG and grown o/n for 16 h.

The cells were harvested, re-suspended in lysis buffer (50 mM Tris, 150 mM NaCl, 20 mM imidazole, pH 7.5) supplemented with 50 µg Dnase from bovine pancreas (Sigma), 50 µg lysozyme from chicken egg white (Sigma) and 1 EDTA-free protease inhibitor cocktail tablet (Roche), before being sonicated and centrifuged (50,000 g, 30 min, 277 K on BECKMAN Avanti-J25). The filtered supernatant was applied on a 5 ml HisTrap HP chelating column, washed with lysis buffer, and the protein was eluted with a linear imidazole gradient from 96% buffer A (50 mM Tris, 150 mM NaCl, pH 7.5) to 100% buffer B (50 mM Tris, 150 mM NaCl, 500 mM imidazole, pH 7.5) using an ÄKTA-purifier.

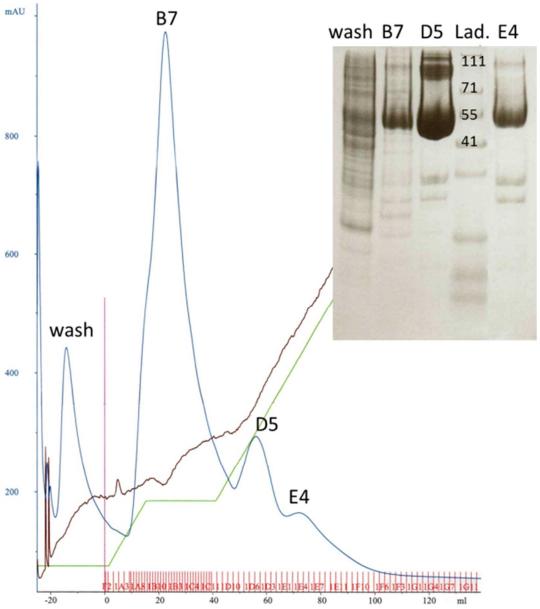


Figure 3.3 Chromatogram of *Tb*UGP (construct B) on a HisTrap HP 5 ml column
The chromatogram describes the UV absorbance on the y-axis in blue in mAU, in green a gradient of Buffer B, in brown is the conductivity of the solution, in red the collection tubes and the volume in ml on the x-axis . The SDS-page gel on the right shows the following from left to right: wash = flow through HisTrap at 4% of buffer B and 96% buffer A; samples from tubes B7, D5; Lad. = protein ladder in kDa (SeeBlue); sample from tube E4. Protein was collected from A6-E5 for further purification.

Fractions containing *Tb*UGP were pooled, concentrated and applied on a Superdex75 16/60 column, equilibrated with buffer C (50 mM Tris, 150 mM NaCl, 10% glycerol, pH 7.5). The size-exclusion chromatography fractions containing *Tb*UGP were pooled and flash frozen in liquid nitrogen. Aliquots were stored at -80°C at a concentration of 2 mg/ml.

Table 3.3 Amino acid sequence of cleaved *Tb*UGP crystallisation construct B

LNPPSAFSGAALACLEKMQASGVEEKCIHIFLIQHALVRKGETGYIPEKSISPVESLPFLQGIETKG
ENTALLRQAVVLKLNGGLGTGMGLNGPKSLLQVKNGQTFLDFTALQLEHFRQVRNCNVPFML
MNSFSTSGETKNFLRKYPTLYEVFDSDIELMQNRVPKIRQDNFFPVTYEADPTCEWVPPGHGD
VYTVLYSSGKLDYLLGKGYRYMFISNGDNLGATLDVRLLDYMHEKQLGFLMEVCRRTESDKKG
GHLAYKDVIDETTGQTRRRFVLRESAQCPKEDEDSFQNIAKHCFFNTNNIWINLMELKKMMD
EQLGVLRLPVMRNPKTVNPQDSQSTKVYQLEVAMGAAISLFDRSEAVVVPRERFAPVKTCSDL
LALRSDAYQVTEDQRLVLCEERNGKPPAIDLDGEHYKMIDGFEKLVKGGVPSLRQCTSLTVRGL
VEFGADVSVRGNVVIKNLKEEPLIIGSGRVLDNEVV

Table 3.4 Biochemical properties of *Tb*UGP crystallisation construct B

Number of amino acids: 483

Molecular weight: 54191.2

Theoretical pI: 6.03

Abs 0.1% (=1g/l): 0.588 at 280nm (Artimo*, et al.,* 2012)

The crystallization construct B was provided by the Hui group (Structural Genomics Consortium, University of Toronto). The construct was transformed and overexpressed in *E. coli* Bl21(DE3) cells as described previously (Marino*, et al.,* 2010). In brief, cells were grown o/n at 37 °C in 10 ml of LB media with ampicillin (50 µg/ml) and transferred into 50 ml of TB media + ampicillin (50 µg/ml) and incubated for 3 h at 37 °C. Subsequently the culture was transferred into 1.8 L of TB supplemented with ampicillin and grown till an OD_{600} of 1 was reached.

Overexpression of *Tb*UGP was induced by adding 0.5 mM for an overnight incubation at 15 °C. For harvest and protein purification the protocol described above (3.1.3) was used and adapted with following changes. The lysis buffer D used was 50 mM HEPES pH 7.5, 500 mM NaCl, 5 mM imidazole, and 5% glycerol, and the wash buffer E used was 50 mM HEPES pH 7.5, 500 mM NaCl, 30 mM imidazole, and 5% glycerol. After the HisTrap-column, the protein was treated with TEV protease (produced in-house) and dialysed into 10 mM HEPES, 500 mM NaCl, 5 mM imidazole and 5 mM DTT o/n. After dialysis, the protein was again applied on a HisTrap -column to separate the TEV – tagged and untagged *Tb*UGP (Figure 3.4.).

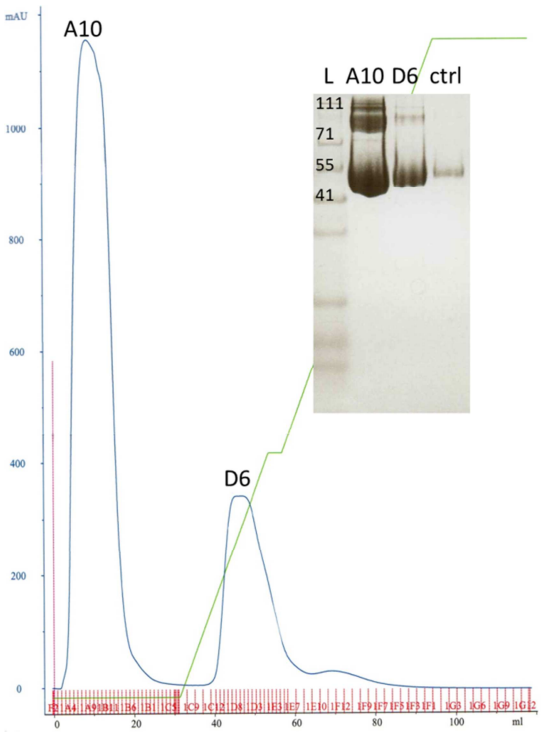


Figure 3.4 Chromatogram of *Tb*UGP (construct B) on a HisTrap HP 5 ml column after TEV – cleavage

The chromatogram describes the UV absorbance in blue on the y-axis in mAU and volume on the x-axis in ml, in green a gradient of buffer B, in brown is the conductivity of the solution and in red the collection tubes. The SDS-page gel on

the right shows the following from left to right: L = protein ladder in kDa (SeeBlue); A10 = flow through HisTrap; D6 = uncleaved *Tb*UGP; ctrl = purified *Tb*UGP from previous purification.

The eluation buffer for Superdex75 16/60 column was 10 mM HEPES, pH 7.5 and

500 mM NaCl, 5 mM DTT.

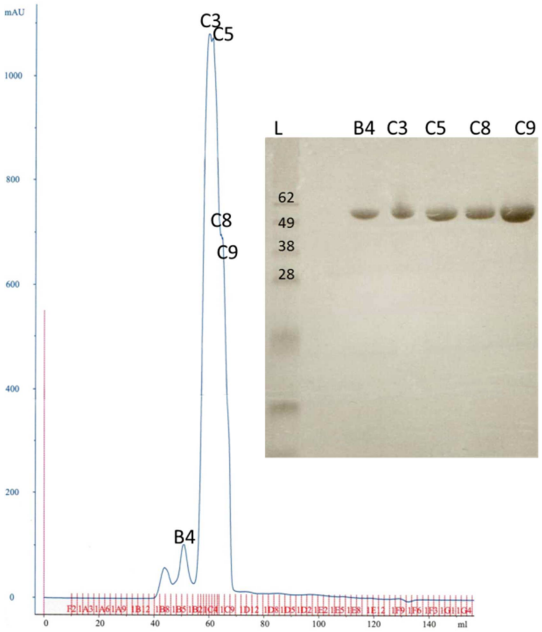


Figure 3.5 Chromatogram of *Tb*UGP (construct B) on a Superdex75column after TEV – cleavage

The chromatogram describes the UV absorbance on the y-axis in blue in mAU, volume on the x-axis in ml and in red the collection tubes. The SDS-page gel on the right shows the following from left to right: L = protein ladder in kDa (SeeBlue); samples from the tubes B4, C3, C5, C8 and C9. Protein from C3 to C9 was collected for further experiments.

The protein was concentrated up to 17 mg/ml and aliquots were flash frozen in

liquid nitrogen and subsequently stored at -20°C.

3.1.4 Kinetic inhibition assay of *Tb*UGP

Inhibition of *Tb*UGP was measured using a pyrophosphatase coupled Biomol Green assay (Enzo Life Sciences) (Figure 3.6) based on previously published paper (Marino, *et al.,* 2010). Biomol Green is a modified malachite-green and therefore very sensitive in detecting free phosphate in solution. The change of colour from yellow to green was detected at 650 nm in a 96 well plate with a SPECTRA max 340 PC spectrometer.

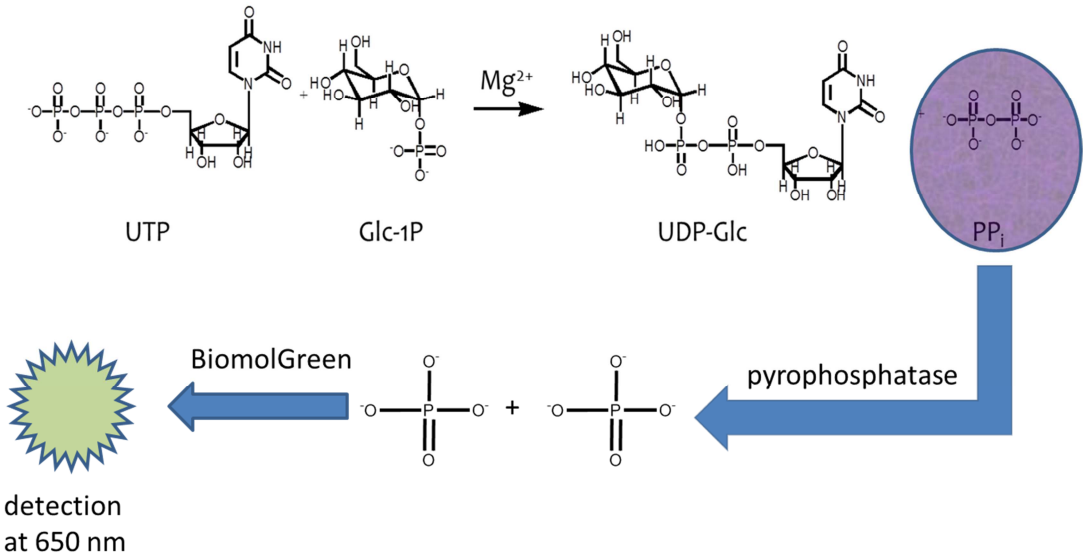


Figure 3.6 Detection of phosphate using BiomolGreen
Reaction mechanism of *Tb*UGP. The produced pyrophosphate (PPi) is converted by pyrophosphatase into free phosphate which is then coloured by BiomolGreen and detected in a plate reader at 650 nm.

The DMSO tolerance for *Tb*UGP in this assay was determined to 5%. The aqueous buffer consisted of 50 mM Tris/HCl pH 7.2, 1 mM DTT, 0.1 mM EDTA, 0.1 mM EGTA, 150 mM NaCl, 10 mM $MgCl_2$ and 0.02% CHAPS. The substrate concentration was at K_m: 122 µM Glc1P and 53 µM UTP. At a *Tb*UGP concentration of 30 fM the reaction

was linear for 30 min. Pyrophosphatase concentration was at 0.1 units per ml. The reaction was stopped after 30 min by adding the same volume as the reaction volume of Biomol Green. After a colour developing time of 30 min the absorption was measured at 650 nm. All compounds were tested with LCMS for purity and in the biochemical assay as well in absence of *Tb*UGP to make sure, that inhibitors found are not inhibiting pyrophosphatase instead of *Tb*UGP. As a standard inhibitor non-hydrolysable Rp-UTP-α-S (BioLog, Germany) was used.

3.1.5 High throughput screening (HTS) with *Tb*UGP

The HTS was carried out together with Raffaella Grimaldi. The assay conditions were the same as described in 3.1.4., except that different instruments (see below) and plates (384 PSF clear well plates) were used. The in-house DDU compound sets consisted of a general set (62538 compounds, (Brenk*, et al.,* 2008)), a kinase set (6724 compounds, (Brenk*, et al.,* 2008)), the Prestwick Library (1120 compounds) and of a fragment library (652 compounds). All HTS libraries were tested at 35 µM in the primary screen (single point (SP) and hits from SP in double point measurements (DP)) and at 100 µM top concentrations in the potency screen (ten point). The fragment set was screened at 500 µM. Compounds were transferred from the source plates to the screening plates using the HummingBird instrument (DIGILAB). The substrate mix (UTP, Glc1P and pyrophosphatase in buffer) was then pipetted into the plates and the reaction was started by adding substrate using an

8-channel micro plate dispenser (Thermo Scientific Matrix WellMate). In addition, Biomol Green was added and assay plates were read using a PerkinElmer 2102 Multilabel Reader (Envision), which can read several plates automatically. To monitor assay performance, each dispensing and measuring instrument was tested before the screen and calibrated if necessary. Two control rows (high/low) were measured on each plate together with a standard inhibitor **16** (Table 3.6) or UTP-α-S (**7**, Table 3.5). The high-control was the reaction mixture without compounds and the low control was the complete reaction mixture without *Tb*UGP. The signals of these wells were averaged (AV_{HIGH}, AV_{LOW}) and used to calculate the signal to noise ratio (AV_{HIGH}/AV_{LOW}) and

$$Z' = 1 - \frac{3(SD_{HIGH} + SD_{LOW})}{AV_{HIGH} - AV_{LOW}}$$

Equation 3.1 Z-factor (Zhang*, et al.,* 1999)

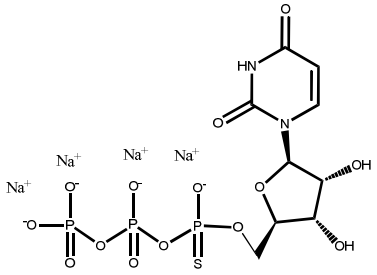
(SD: standard deviation; Z´: Z-factor). The ligand efficiency was calculated using the following formula:

$$LE = \frac{-RTlnK_i}{number\ of\ heavy\ atoms}$$

Equation 3.2 Ligand efficiency

(LE = ligand efficiency in kcal/mol per non-hydrogen atom; R = ideal gas constant; T = 298K, K_i = binding affinity calculated from IC_{50} values using

Table 3.5 Non-hydrolysable substrate UTP-α-S

#	Structure	pIC_{50} for WT *Tb*UGP (G219I– mutant)	Hill slope
7		4.51±0.02; (4.86±0.01)	0.9±0.1 (0.8±0.1)

3.1.5.1 Analysis of HTS hits to establish SAR

In order to be able to establish a SAR, the software ClassPharmer was used with the following settings to cluster the screening hits:

- Cluster by ring system with minimum 1 ring and maximum 3 rings
- Skip compounds which have more than 20 rings
- Do not allow portioning of fused rings into individual rings
- Do not allow connecting of rings using non-ring linker atoms
- Set in class parameters that the homogeneity level is high, redundancy level is low.

Compounds spanning a range of potencies in each cluster were selected using Pipeline Pilot (Accelrys). Pipeline Pilot is a graphical workflow application, where different modules can be used to process, analyse, manipulate and filter tables, texts and graphics. In the first step all singletons were rejected. To be able to probe one cluster for active compounds, three compounds from each cluster were chosen according to the following rules:

- From each cluster take the compound with the highest PI
- Take a compound from the same cluster with PI closest to >70% inhibition
- Take a third compound X (if present) from the same cluster for which PI is over 70% -inhibition and less than highest PI.

3.1.6 NMR fragment screen and analysis

The NMR fragment screen was carried out as described for *Gs*6PGDH enzyme (2.1.2.1 and 2.1.1) except that for competition 5 mM UTP was added.

3.1.7 Testing solubility using a nephelometer

A nephelometer detects insoluble particles in solutions by measuring forward scattered light. The particle density in the solution is a function of the reflected light hitting the detector. Because the amount of light reflected depends on the shape and reflectivity of the particles, a reference sample must be measured to calibrate

the nephelometer. Nephelometry was used to determine solubility of compound **16**. As a reference the assay buffer (3.1.4) plus DMSO (2%-5%) was used for calibration. Compound **16** was added to the same solution at different concentrations (0.1 mM – 3 mM) and measured in triplicates. The reflectivity is giving a numeric value. If this value is not more than three times higher as the correspondent reference value, the compound was considered to be soluble at the given concentration.

3.1.8 DNA isolation and manipulation

Plasmid DNA was purified from *E. coli* (DH5α) using the Miniprep kit from Qiagen. The mutation G219I was inserted into the crystallisation construct using site-directed mutagenesis (Stratagene QuikChange). The following primers (Eurofins MWG Operon in Germany) were used plus 10% DMSO in the reaction mix:

Forward: 5’ CTACATGTTTATATCAAACATAGACAACCTTGGCGCGAC 3’

Reverse: 5’ GTCGCGCCAAGGTTGTCTATGTTTGATATAAACATGTAG 3’

The DNA was sequenced by the University of Dundee oligonucleotide facility using T7 (TAA-TAC-GAC-TCA-CTA-TAG-GG) + T7 terminator (CTA-GTT-ATT-GCT-CAG-CGG-TG) primers.

3.1.9 Crystallisation of *Tb*UGP G219I-mutant and structure determination

Crystals were grown using the hanging drop method in 24-well plates on VDX slides (Hampton Research) or with the Phoenix RE (Art Robbins Instruments) using the sitting drop method in 96-well MRC 2 crystallisation plates. Several screening sets from Hampton Research (PEG/Ion, Index) and from Qiagen (Classics, JCSG+, $AmSO_4$ Suite) were screened for suitable crystallisation conditions using a robotic system Phoenix RE (Art Robbins Instruments).

Hanging drop method

The G219I-mutant was crystallised in presence of 3 mM UDP-Glc under the same conditions as described earlier (Marino*, et al.,* 2010). Crystals were grown using 1 µl crystallisation buffer at pH 5.5 (22% PEG 3350 (as precipitant), 0.1 M ammonium sulphate, 0.1 M BisTris) from the reservoir and mixed with the same volume of 17 mg/ml G219I-mutant (in the following buffer: 10 mM HEPES, pH 7.5 and 500 mM NaCl and 5 mM DTT) and equilibrated against 500 µl of reservoir. Crystals appeared after growing for two days at 20 °C. These were flash-cooled in crystallisation buffer with 20% PEG 400.

Sitting drop method

Construct B of *Tb*UGP (Table 3.3) was crystallised with a 100 nl mixture of 3 mM **16** and 17 mg/ml *Tb*UGP (in the same buffer as G219I-mutant described above) together with either 100 nl or 200 nl of the following crystallisation buffer: 0.2 M lithium sulphate, 0.1 M Tris/HCl pH 8.5 and 1.26 M ammonium sulphate as precipitant. Crystals appeared after growing for two days at 20 °C. These were flash-cooled in crystallisation buffer with 20% PEG 400 before collection.

3.1.9.1 Data collection, processing and structure modelling

X-ray diffraction data for the mutated *Tb*UGP were remotely collected at beamline ID14 in the European Synchrotron Radiation Facility (ESRF) in Grenoble (France), equipped with a MAR 225 CCD detector. A full data set of 200 images were collected from 140 ° - 340 ° with ϕ = 1 and 4 s exposure time. Crystal orientation, cell parameters and possible space group were determined using MOSFILM (Leslie and Powell 2007; Battye*, et al.,* 2011). Generated reflection lists and integrated reflections from the images were scaled and merged using SCALA (Evans 2006; 2011) from the CCP4 suite of programs (Potterton*, et al.,* 2003). Resolution data less than 45 or greater than 2.5 Å were excluded to produce an R_{merge} of 0.121.

Model generation was done using PHASER (McCoy 2007; McCoy*, et al.,* 2007). As a model enzyme the coordinates from *Tb*UGP (PDB 3GUE) dimer were used as a starting point with a sequence identity set to 0.95 and molecular weight to 56 kD. A

solution was found with a rotation function score (RFZ) = 46.0 and a translation function score (TFZ) = 36.1. The structure was refined using REFMAC5 (Murshudov*, et al.,* 2011) from the CCP4 package. Refinement was done in 20 cycles using a manual weighting term of 0.1 at first and automatic weighting for the last refinement steps. Iterative model building was carried out using the interactive graphics program WinCOOT (v. 0.7.1 -pre) (Emsley and Cowtan 2004; Emsley*, et al.,* 2010). At first, all amino side chains with no electron density were mutated as stubs. The ligand UDP-Glc was added to the structure using the function "Find Ligands" and refined using "Real Space Refine Zone" feature. After a refinement with REFMAC5 the amino side chains were mutated back and the occupancy of the atoms without electron density were set to 0.01. The structure was refined using the validation steps available in WinCoot and checked using web-based validation server MolProbity (http://molprobity.biochem.duke.edu/) (Chen*, et al.,* 2010) and RCSB PDB (http://validate.rcsb.org/).

3.2 Results

3.2.1 Hit discovery by virtual screening

3.2.1.1 Virtual screening

The crystal structure of the *Tb*UGP –product complex was determined (Marino*, et al.,* 2010).

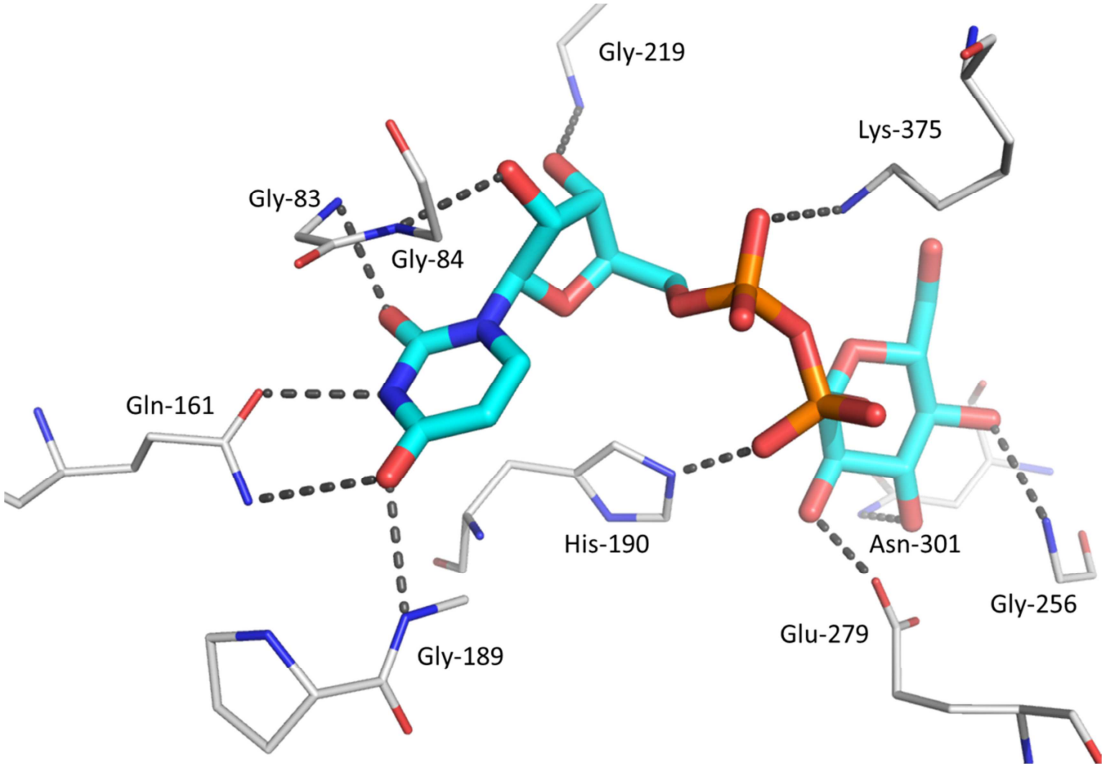


Figure 3.7 Binding mode of the product UDP-Glc in *Tb*UGP
Hydrogen-bond interactions between product and *Tb*UGP are indicated as dotted lines. Nitrogen atoms are coloured blue, oxygen atoms red, phosphate atoms orange, and carbon atoms of the protein white and of the ligand light blue.

The uracil moiety forms two hydrogen bonds to the enzyme backbone (Gly83 and Gly189) and two with the side chain of Gln161 (Figure 3.7). The uridine ring stacks with the amide bonds of Gly83, Gly84 and Gly189. The 2-hydroxy group of the ribose ring points into a cavity which appears large enough to accommodate a mono substituted cyclopentane ring (Figure 3.8a). The phosphate group forms two hydrogen bonds with the Lys375 and His190 side chains and the buried glucose residue forms hydrogen bonds to Glu279, Glu256 and Asn301.

To assess if there is a rational basis for selective inhibition of *Tb*UGP over *h*UGP the binding sites of both enzymes were compared. At the time of study no crystal structure of *h*UGP was available. Therefore, a homology model was build (3.1.1). Comparison of the homology model and the *Tb*UGP crystal structure revealed that all hydrogen-bond interactions that the product forms in *Tb*UGP are conserved in the model structure of the complex (Figure 3.9). However, the cavity lying opposite of the 2-hydroxy group of the ribose ring of the product in *Tb*UGP is not present in the model structure of *h*UGP (Figure 3.8 b)). This is due to a replacement of Gly219 in *Tb*UGP with Ile252 in *h*UGP.

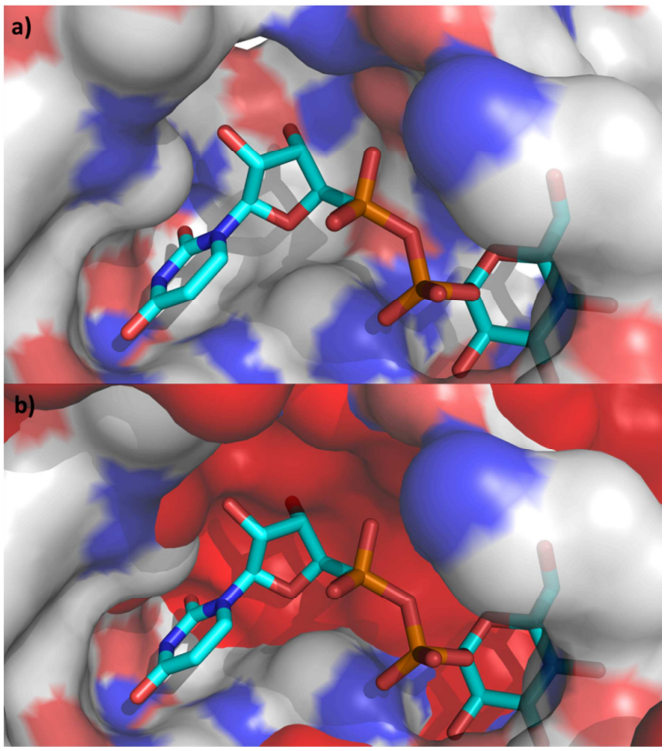


Figure 3.8 Surface overlay of *Tb*UGP with *h*UGP-model (red) with bound UDP-Glc
a) Solvent accessible surface of the binding site of *Tb*UGP. (b) Alignment of *Tb*UPG (coloured atoms) and the *h*UGP-model (red). For clarity, only the solvent accessible surface of the binding sites and the bound UDP-Glc are shown. The cavity opposite the 2-hydroxy group of ribose ring (selectivity pocket) is blocked in the human model.

It can be speculated that ligands occupying this cavity would bind selectively to *Tb*UGP. From this point onwards, this cavity is called the selectivity pocket.

Recently, the apo-structure of *h*UGP was published (Yu and Zheng 2012). Comparison of homology model and this structure (PDB code 3R2W) revealed that the build *h*UGP model is structurally very close to the published X-ray structure (aligned Å, Figure 3.9). The side chain location of Ile241 in the *h*UGP X-ray structure blocking access to the selectivity pocket was also predicted correctly (Figure 3.10).

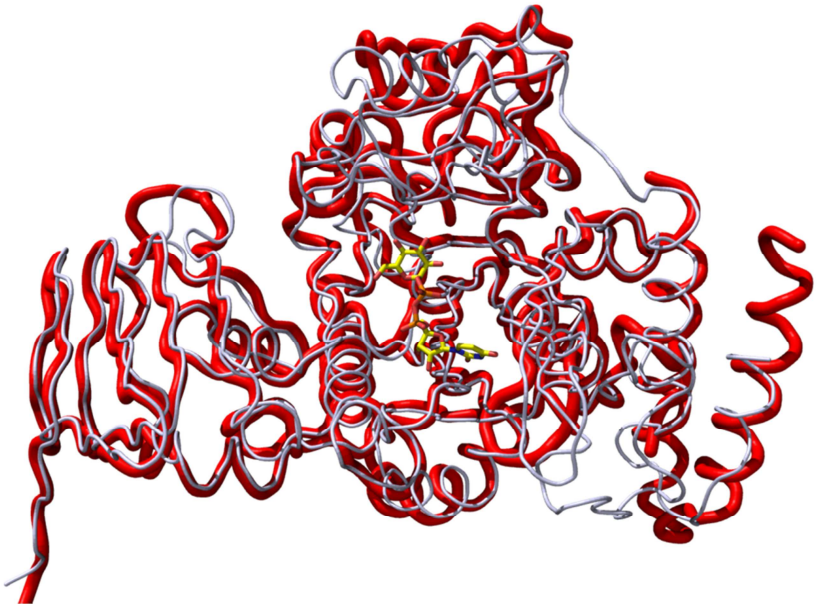


Figure 3.9 C-alpha chains overlay of *h*UGP(red) and *h*UGP-homology model (white)
C-alpha chains of *h* UDP-Glc from *Tb*UGP (yellow carbon atoms) is displayed to indicate the location of the active site.

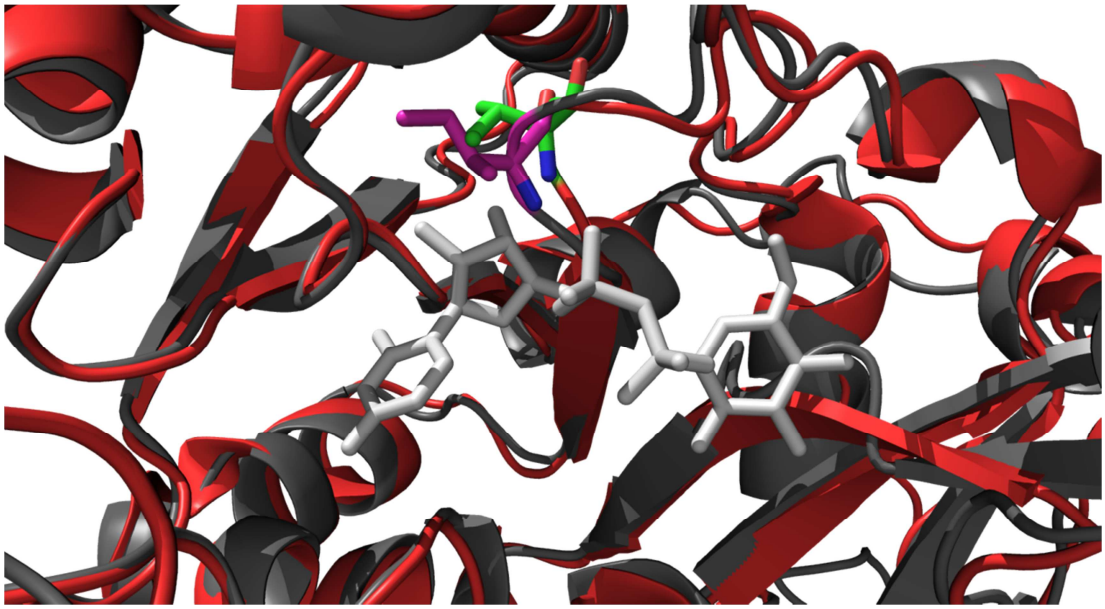


Figure 3.10 Overlay of active sites of *h*UGP(red) and G219I-mutant(black)
UDP-Glc from G219I-mutant (in white) is displayed to indicate the location of the active site. In pink is Ile241 of *h*UGP and in green Ile219 of the G219I mutant.

A hierarchical screening protocol (Figure 3.12) was established to retrieve small compounds that bind into the uracil pocket of *Tb*UGP and possibly extend into the selectivity pocket. First, a database of commercially available compounds was filtered according to physicochemical criteria. Next, a pharmacophore hypothesis was derived and used to filter all compounds passing through the first filter. The remaining compounds were docked into the *Tb*UGP binding site and promising compounds with a favourable, predicted binding mode, were purchased.

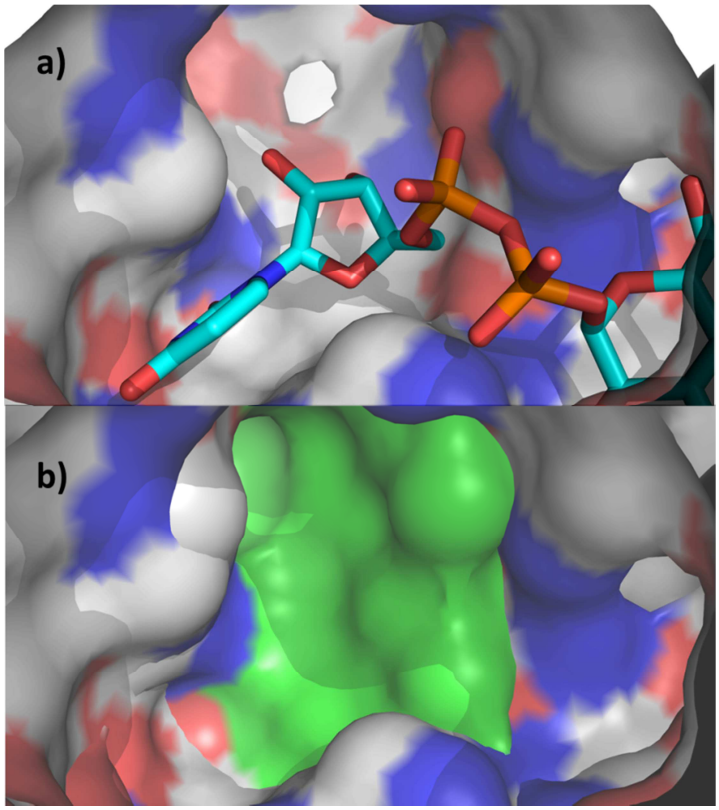


Figure 3.11 Surface overlay of *Tb*UGP with *h*UGP (green) with bound UDP-Glc in *Tb*UGP.

a) Solvent accessible surface of the binding site of *Tb*UGP.
b) Alignment of *Tb*UPG (coloured atoms) and the *h*UGP-crystal structure (green) (PDB 3R2W). For clarity, only the solvent accessible surface of the binding sites and the bound UDP-Glc are shown. The cavity opposite the 2-hydroxy group of ribose ring (selectivity pocket) is blocked in the human structure, which is consistent with the human model (Figure 3.8).

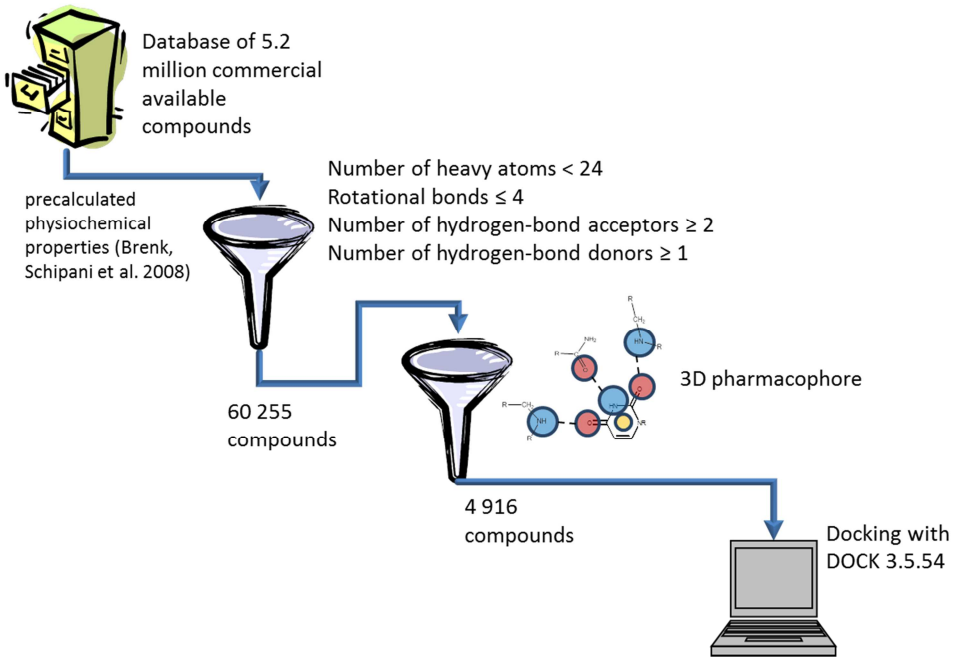


Figure 3.12 Virtual screening cascade used to identify potential *Tb*UGP inhibitors

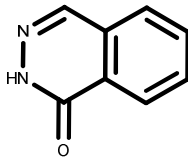
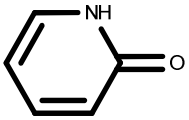
An in-house virtual database of 5.2 million commercially available compounds (Brenk*, et al.,* 2008) was filtered according to the following criteria to derive a compound set for virtual screening: Only compounds that had less than 24 heavy atoms, four or less rotational bonds, two or more hydrogen-bond acceptors and one or more hydrogen bond donors, no unwanted (toxic or reactive) groups, were allowed to pass this filter. The selection criteria were chosen to obtain fragment-like compounds binding into the uracil pocket which were big enough to reach the selectivity pocket. In total, 60255 compounds passed this filter.

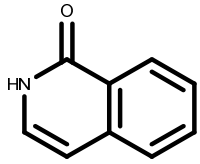
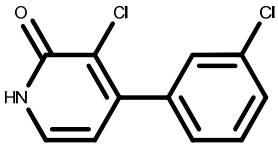
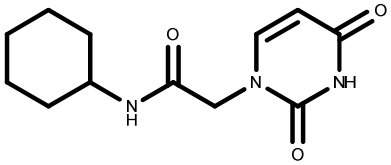
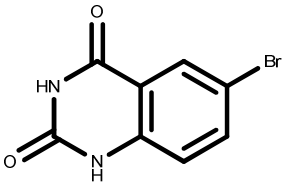
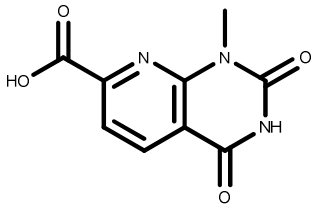
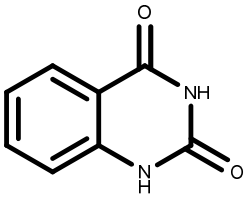
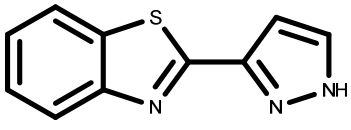
The selected subset was further filtered using a protein-based 3D pharmacophore to select for compounds that have the required spatial arrangement of functional groups to bind into the uracil pocket. The pharmacophore (Fig 3.1) consisted of three hydrogen-bond features and an aromatic ring feature to mimic all atoms of

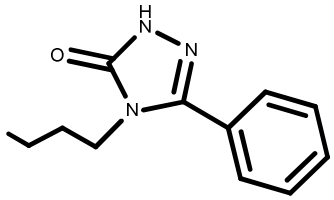
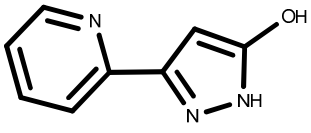
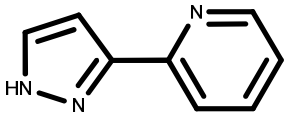
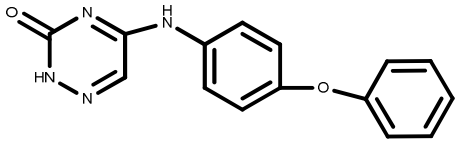
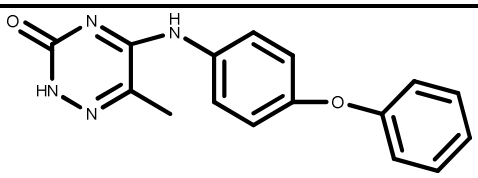
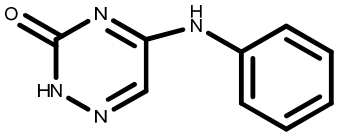
the uracil moiety of the product that from hydrogen-bonds with the protein and the aromatic ring that stacks with amide bonds (Fig 3.3). With this setup 4916 compounds fulfilled the pharmacophore requirements.

In the final step, the compounds passing the previous filter steps were docked into the receptor. For all of these compounds a binding mode was generated. Visual analysis of the docking poses revealed that many compounds did not occupy the uracil binding pocket. As interactions in this pocket were judged to be crucial for binding, these compounds were not further considered. In total, 15 of these remaining compounds were purchased for binding studies (Table 3.6). Out of those four were ring fragments of docked compounds (**8**, **9**, **10**, **15**) that appeared frequently in high-ranking compounds.

Table 3.6 Compounds shortlisted for binding assays with *Tb*UGP after virtual screening

#	Structure	pIC_{50} for *Tb*UGP	Hill slope
8		not active	not active
9		not active	not active

#	Structure	pIC_{50} for *Tb*UGP	Hill slope
10		not active	not active
11		insoluble	
12		not active	not active
13		insoluble	
14		not active	not active
15		not active	not active
16		3.53±0.04	1.1±0.1

#	Structure	pIC_{50} for *Tb*UGP	Hill slope
17		not active	not active
18		not active	not active
19		not active	not active
20		insoluble	
21		insoluble	
22		3	

3.2.1.2 Evaluation of virtual screening hits

Initial solubility tests indicated that compounds **11**, **13**, **20** and **21** were not soluble at the required concentration for compound evaluation. Therefore these compounds were omitted in further studies.

3.2.1.3 Binding studies using NMR

NMR provides a relatively cheap, fast and sensitive method to measure weakly binding compounds up to mM range (Meyer 1999). With this technique it is also possible to identify a binder out of a mixture of several compounds. Therefore, compounds were tested for binding to *Tb*UGP using two NMR experiments. First, ^{1}H-NMR spectra for all compounds were recorded. Subsequently, cocktails containing a mixture of 10 different compounds with the protein were prepared and STD (Figure 3.13, Figure 3.15) and wLOGSY (Figure 3.14, Figure 3.16) spectra were measured. The compounds which gave a positive signal in the STD and wLOGSY experiments were tested for interactions. The substrate UTP was added to the mixture to displace bound compounds in the active site and the spectral measurements were repeated. Only two compounds (**16**, **22**) showed binding to *Tb*UGP as in both experiments (STD + wLOGSY, Figure 3.14 and Figure 3.15). A signal reduction after adding the substrate UTP was only observed for compound **16** (Figure 3.14.) indicating that this compound was accommodated in the UTP binding site.

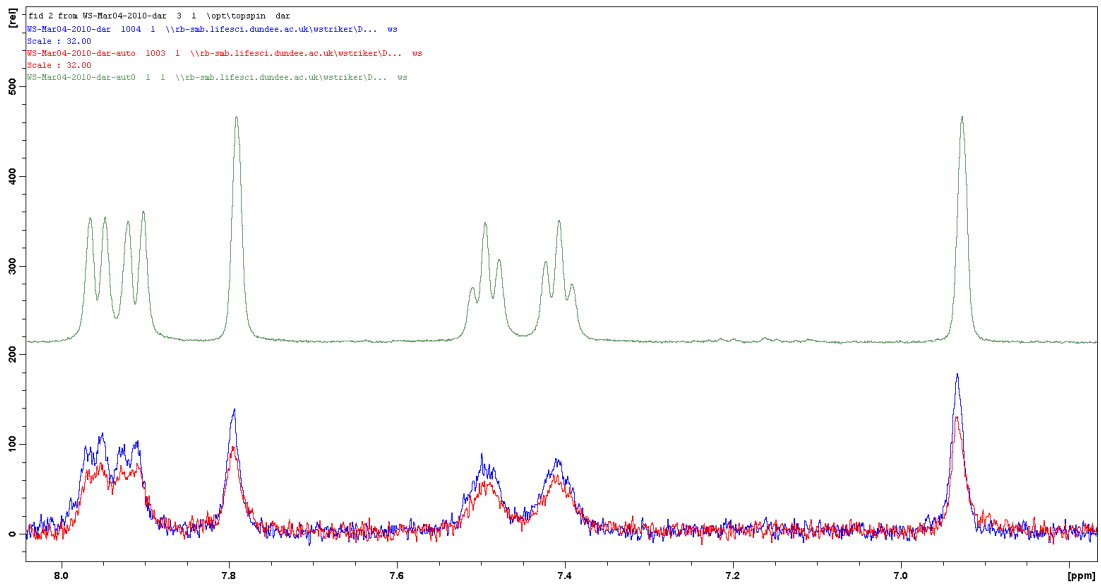


Figure 3.13 NMR STD - spectra for 16 showing signal reduction after adding UTP
In green is the reference spectrum of compound **16**, in blue the STD – spectrum of the compound and in red the STD – spectrum when substrate UTP is added.

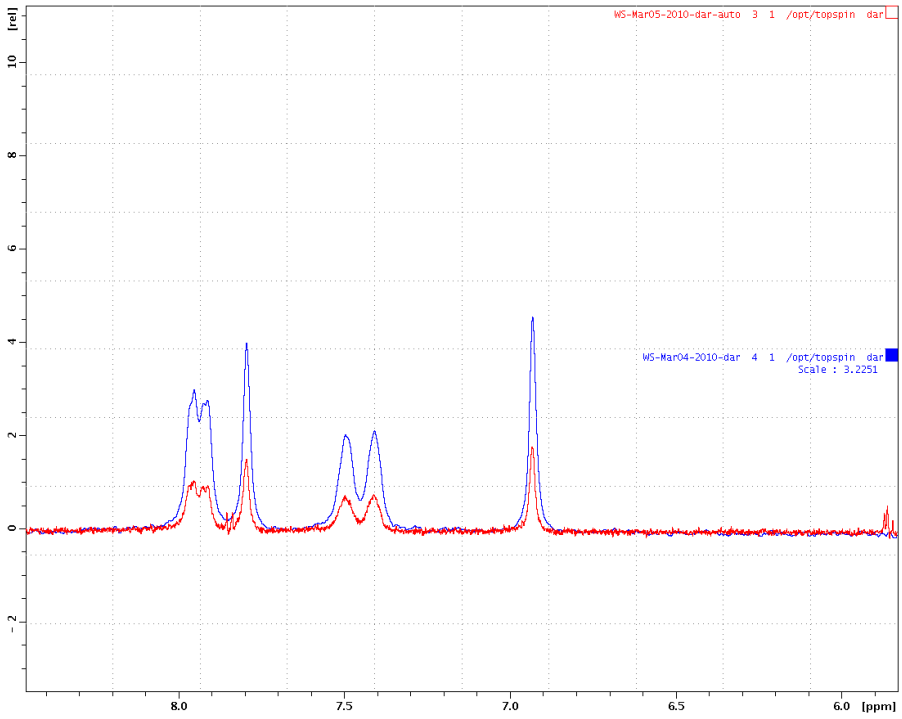


Figure 3.14 NMR wLOGSY - spectra for 16 showing signal reduction after adding UTP

A wLOGSY spectrum from *Tb*UGP with inhibitor **16** added. In blue is the wLOGSY spectrum without the substrate UTP and in red the reduced signal when UTP is added.

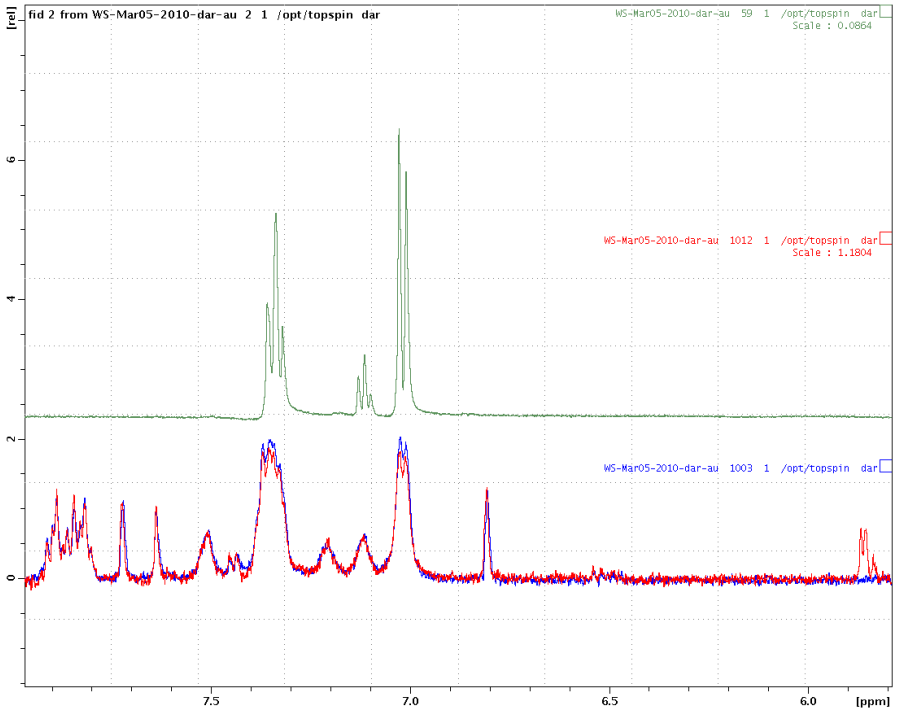


Figure 3.15 *Tb*UGP - STD spectra of compound 22 in a mixture
In green is the reference spectrum of compound **22**, in blue the STD – spectrum of the compound mixture and in red the STD – spectrum when substrate UTP is added.

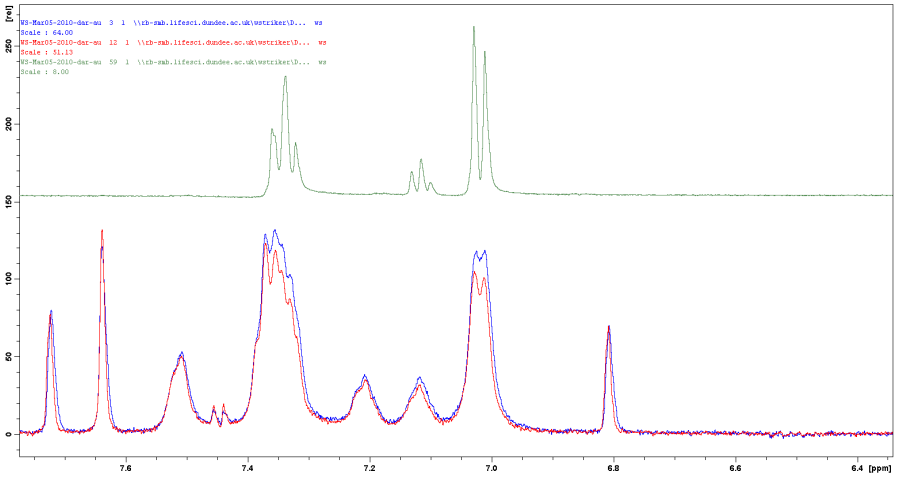


Figure 3.16 *Tb*UGP - wLOGSY spectra of compound 22 in a mixture

In green is the reference spectrum of compound **22**, in blue the wLOGSY – spectrum of the compound mixture and in red the wLOGSY – spectrum when substrate UTP is added.

3.2.1.4 Inhibition assay with virtual screening hits

All compounds (Table 3.6) were tested in two independent measurements. For each measurement one replication was carried out and was performed under the same conditions as previously described (3.1.5) at concentrations up to 1 mM (as described in 3.1.5). The only inhibitors found were **16** and **22**, with **16** showing 76% inhibition at 1 mM and **22** showing 50%. Only for compound **16** a dose-response curve could be recorded. To ensure that the compound was dissolved at high concentrations, the solubility was determined using a nephelometer. The data showed that compound **16** was not entirely soluble at 2 mM when the assay buffer contained 1 % DMSO. However, when the DMSO concentration was increased to 5 % compound **16** was completely dissolved. Therefore, a DMSO concentration of 5 % was used in the following assays. Under these conditions, the inhibitor has a pIC_{50} of 3.53±0.04 and a Hill slope of 1.1±0.1 (Figure 3.17).

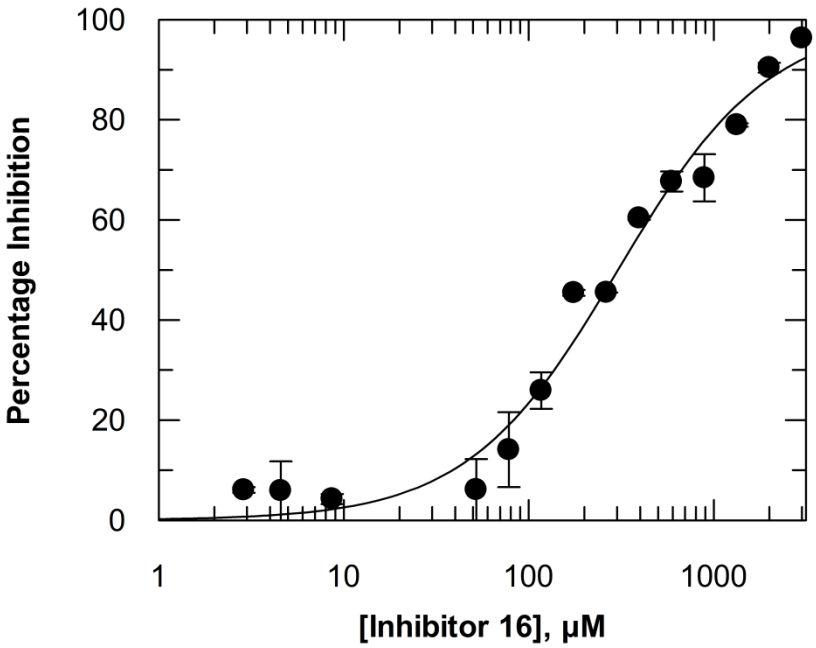


Figure 3.17 Dose response curve of compound 16
Dose-response curve with error bars for each data point from two experiments (each with n=4), percentage inhibition on y-axis and compound concentration in µM on x-axis (logarithmic scale). Points were fitted using a two parameter equation (Equation 2.3). Analysis from two experiments returns a pIC_{50} = 3.53±0.04 and a Hill slope of 1.1±0.1 (mean ± SD; N=2).

3.2.1.4.1 Determination of mode of inhibition of compound 16

Inhibition of **16** was measured at different substrate (UTP) concentrations to determine the mode of inhibition of this compound. Following the Cheng-Prusoff equation

$$IC_{50} = K_i\left(1 + \frac{[S]}{[K_m]}\right)$$

Equation 3.3 Cheng – Prusoff equation

(Cheng and Prusoff 1973) the IC_{50} for a competitive inhibitor rises with increasing concentration of the competing substrate is increased. With a substrate concentration at K_m the $IC_{50} = 2K_m$. When substrate concentration is increased to 5 times K_m IC_{50} will increase 3 times (from 303 µM to 909 µM; $IC_{50} = 6K_m$) and an increase of the substrate concentration to 10 times K_m will increase IC_{50} 5.5 times (from 303 µM to 1667 µM; $IC_{50} = 11K_m$).

Therefore, in the inhibition assay the UTP concentration was increased to 5 times and 10 times the K_m value of UTP (53 µM) to determine if the IC_{50} values changed as expected for a competitive inhibitor. Under these conditions, 29% and 11% enzyme inhibition at 2 mM of **16** were observed respectively (Figure 3.18). Due to solubility limits of **16** a full dose- response curve could not be measured and no IC_{50} values at higher substrate concentration could be determined. Nevertheless, a clear loss of activity could be observed as expected for a competitive inhibitor.

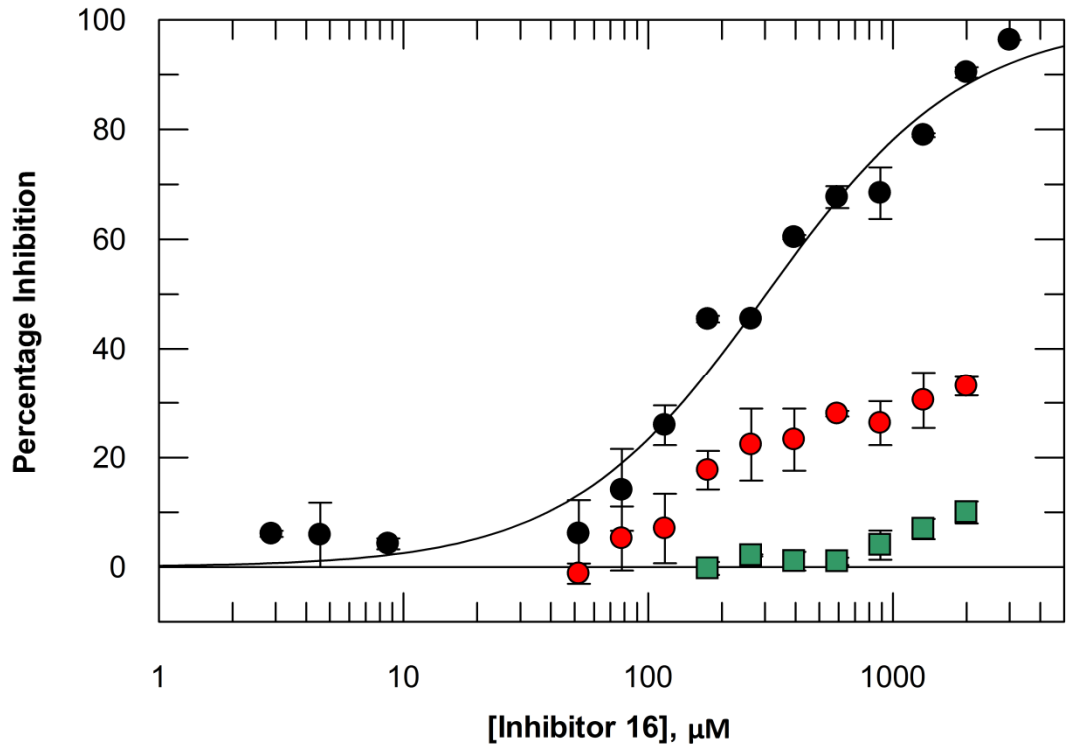


Figure 3.18 Shift in pIC_{50} for 16 at a substrate concentration of 5 times and 10 times the K_m
Dose response curve of **16** (black dots) was taken from (Figure 3.17). Red dots were determined at 5 times K_m (265 µM) and green squares at 10 times the K_m (530 µM). Percentage inhibition on y-axis is plotted versus compound concentration in µM on x-axis (logarithmic scale). Points were fitted using a two parameter equation (Equation 2.3). (mean ± SD; N=2).

3.2.1.4.2 Evaluation of binding mode of compound 16

An attempt was made to determine the binding mode of **16** in *Tb*UGP using X-ray crystallography. The strategy was either to obtain apo crystals of *Tb*UGP to be able to soak compound **16** or to co-crystallise *Tb*UGP with **16**. The best co-crystallised crystals grew in the JCSG+ set (position E4) from Qiagen (for crystallisation condition see 3.1.9). Although the crystals were very small (< 20 µM), an attempt was made to collect reflection data at ESRF facility in France. Unfortunately the

collected x-ray data was of bad quality so it was not possible to determine the structure. Attempts to grow larger crystals were unsuccessful.

In the proposed binding mode, part of the ligand is occupying the selectivity pocket which is not present in *h*UGP (Figure 3.8 and Figure 3.19a). For this reason a point mutation was introduced in *Tb*UGP where Gly219 was changed to Ile that is present in *h*UGP at this place and restricts access to the selectivity pocket (Figure 3.19b). We hypothesised, that compound **16** should not fit in the cavity any longer and therefore not bind to the mutated enzyme if the proposed binding mode is correct (Figure 3.19b).

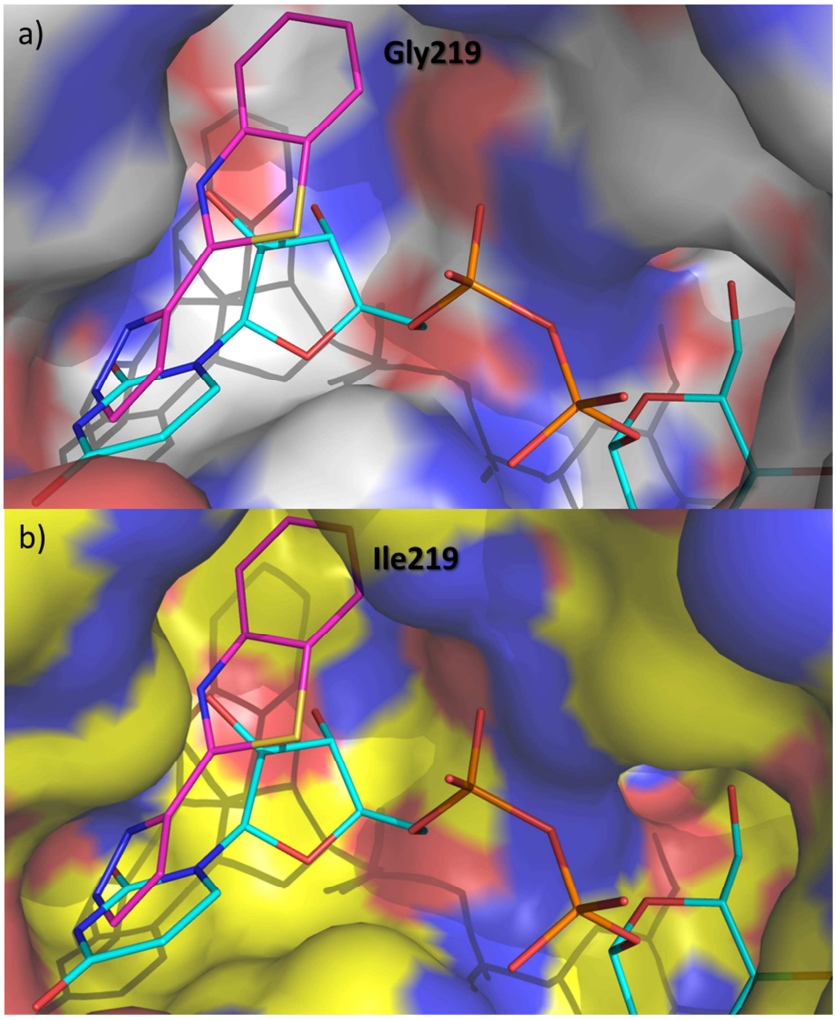


Figure 3.19 Comparison of docked compound 16 in wild type (a) and mutated (b) *Tb*UGP structure
a) Structure of wild type *Tb*UGP with bound product UDP-Glc in blue and proposed binding mode of **16** in pink. The coupled ring system is pointing into the selectivity cavity. b) Structure of the G213I-mutant, where the selectivity cavity is partly closed resulting in a clash with compound **16** in the docked binding mode of the wild type enzyme.

The protein containing the point mutation was purified like the wild type and crystallized under similar conditions (Table 6.2).

In the F_0-F_c electron density map the position of the isoleucine side chain was clearly defined (Figure 3.20). Superposition of G213I from *Tb*UPG and *h*UGP show

that I213 adopts a different conformation (Figure 3.10). As a consequence the selectivity pocket is still partially accessible but not large enough to accommodate **16** in the same binding mode as suggested for the wild type enzyme without a clash (Figure 3.19b).

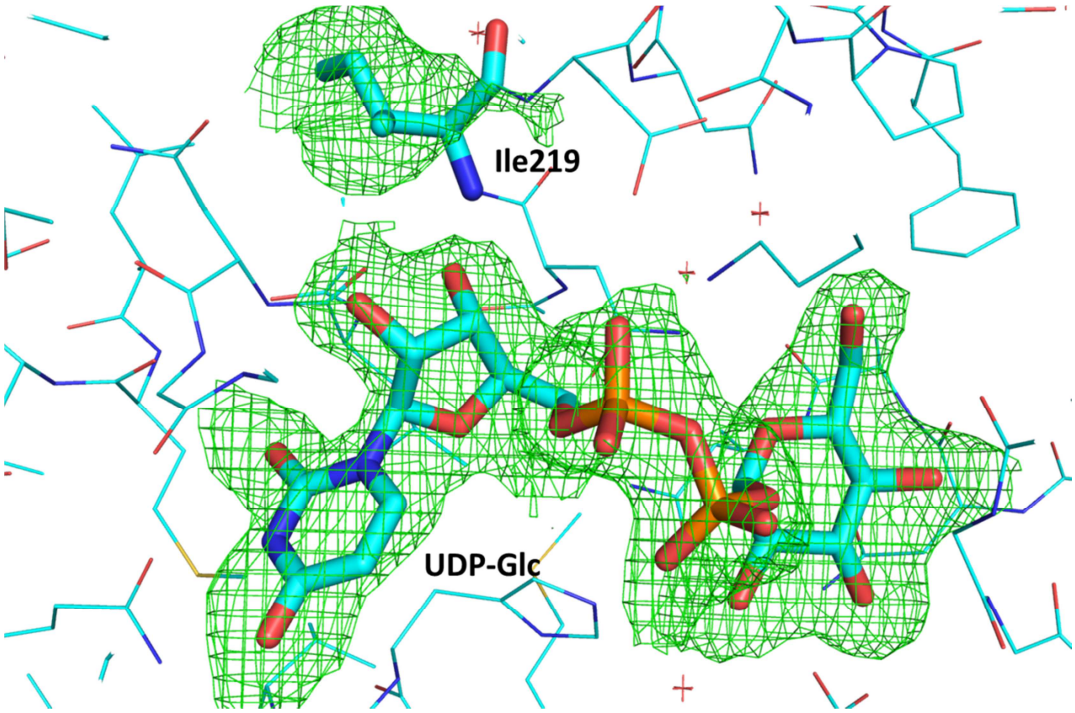


Figure 3.20 Binding site of G219I–mutant with F_O-F_C electron density
The F_O-F_C electron density map shown as green mash (at 1.5 sigma) was calculated without the side chain of Ile219 and UDP-Glc present in the structure.

The mutated enzyme was kinetically characterized. The enzyme was still active but compared to the wild type, the K_m of UTP changed from 53 µM to 18 µM and of Glc-1P from 122 µM to 20 µM (Table 3.7).

Table 3.7 Comparison of K_m values of wild type and G219I-mutant of *Tb*UGP

*Tb*UGP		G219I-mutant	
UTP K_m [µM]	Glc-1P K_m [µM]	UTP K_m [µM]	Glc-1P K_m [µM]
53	122	18	20

Compound **16** showed no activity for G219I-mutant to a concentration up to 1 mM but G219I-mutant was still inhibited by the substrate-like inhibitor UTP-α-S (**7**) (Figure 3.21).

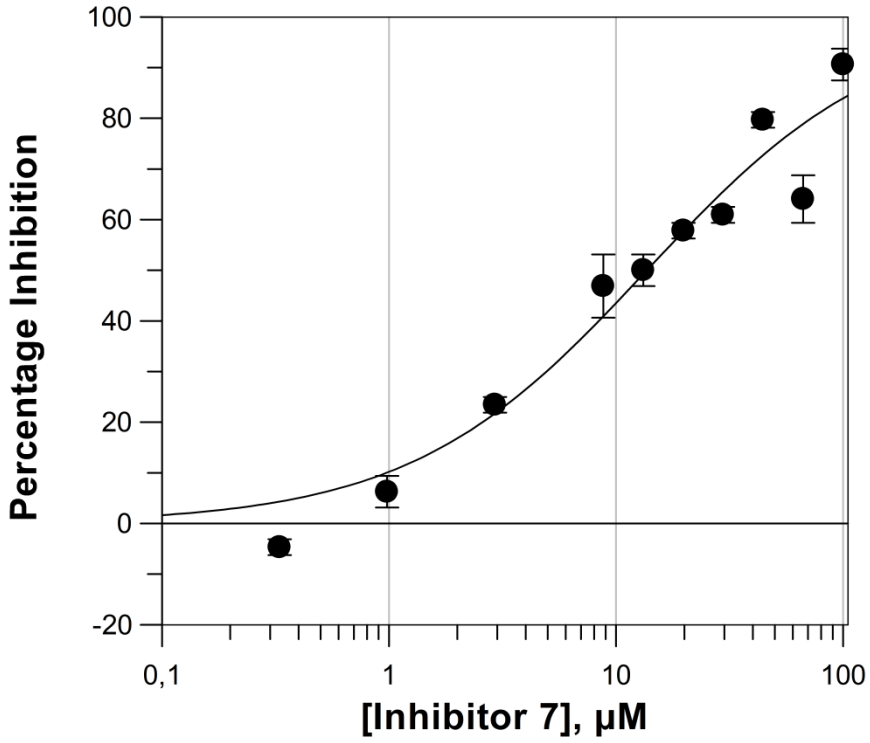


Figure 3.21 Dose-response curve for substrate like inhibitor UTP-α-S
Dose-response curve with error bars for each data point from two experiments (each with n=2), percentage inhibition on y-axis and compound concentration in µM on x-axis (logarithmic scale). Points were fitted using a two parameter

equation (Equation 2.3). Analysis from two experiments returns a pIC_{50} = 4.86±0.01 and a Hill slope of 0.8±0.1 (mean ± SD; N=2).

3.2.1.5 Inhibition of *h*UGP by compound 16

The inhibition constant of **16** with *h*UGP was determined (Figure 3.22). The compound has a pIC_{50} value of 5.3 and a Hill slope of 4.2.

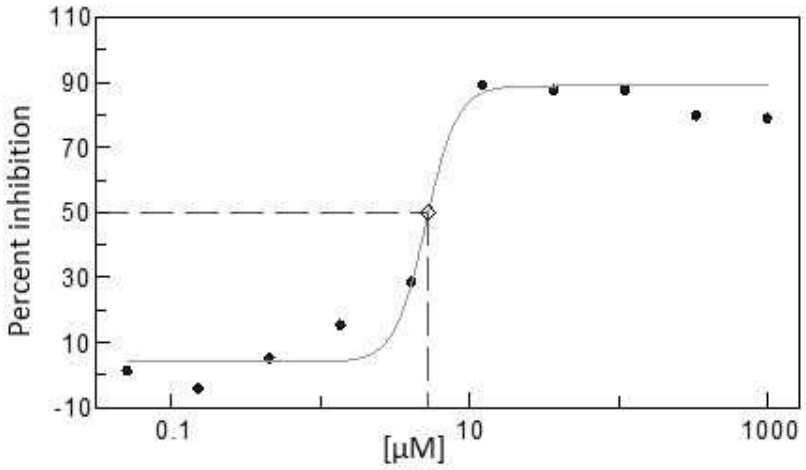


Figure 3.22 Dose response curve of 16 with *h*UGP (measured by Raffaella Grimaldi)
Dose-response curve from one experiment (with n=2), percentage inhibition on y-axis and compound concentration in µM on x-axis (logarithmic scale). Analysis from the experiment returns a pIC_{50} = 5.3 and a Hill slope of 4.2.

Unfortunately, the available amount of *h*UGP was not sufficient to produce any more data with this enzyme.

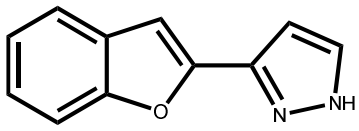
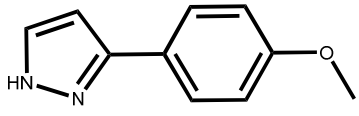
3.2.1.6 SAR around compound 16

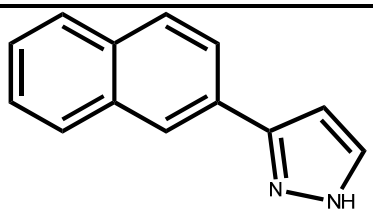
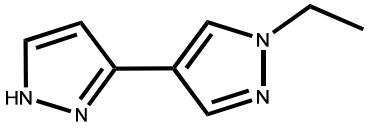
To establish SAR around compound **16**, the Pipeline pilot standard module "Substructure Filter from File" was used to retrieve analogues of compound **16**. This module performed a substructure search in the in-house database (Brenk*, et al.,* 2008). From the few substructures that came through the filter, only 4 were available for purchase (Table 3.8). The four analogues were tested using NMR and biochemical assay.

In the NMR Screen all analogues except **26** gave a signal in the STD and wLOGSY experiments. The signals were reduced when UTP was added suggesting, that they are binding in the substrate binding site.

Unfortunately none of the compounds showed activity up to 1 mM in the BiomolGreen coupled assay.

Table 3.8 Purchased analogues of compound 16

#	*Structure*
23	
24	

#	Structure
25	
26	

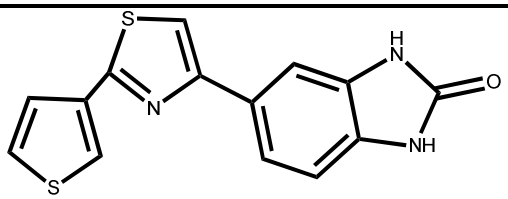
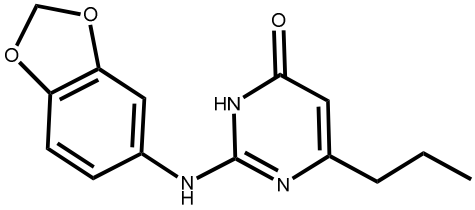
3.2.2 Hit discovery using HTS

The in-house DDU libraries (3.1.5) were screened using BiomolGreen coupled enzyme assay. The best performance was achieved at 30 pM *Tb*UGP and substrate concentration at K_m with a signal to noise ratio = 2.00 and Z´= 0.84 in average.

The HTS primary screen came back with over 2000 hits from the double point – measurements. To see if the hits displayed SAR a maximum substructure analysis was carried out using ClassPharmer™ (Simulations Plus, Inc.). Next, using Pipeline Pilot (Accelrys), three compounds from each generated class were select to cover a range of PI values. This filtering step reduced the amount of compounds to 377. These were analysed by kinetic -potency screens (done by Raffaella Grimaldi) and tested using surface plasmon resonance (SPR) (done by Iva Hopkins Navratilova). The best binders from both screens were compared. The analysis showed that the best hits from the potency screen were generally not confirmed using SPR. Further

the best binders in SPR, except one, had Hill slopes either < 0.5 or > 1.6. Only two compounds (**27** and **28**) were found to be potent in both screens. These compounds do not share the same scaffold or pharmacophore, so no SAR could be developed from it.

Table 3.9 Best binders in SPR and potency screen from filtered HTS

#	Structure	pIC_{50} for *Tb*UGP	Hill slope
27		4.74	1.4
28		5.74	1.1

3.2.3 Screening of fragment library

The fragment library was screened by NMR and by using the kinetic assay. The hit rate of the NMR assay was very high (>40% of screened compounds gave a positive signal in either STD or wLOGSY experiment). However, for most compounds no signal reduction was observed after adding UTP, indicating that the screening setup

is valid. When screening the fragment library with the kinetic assay, a signal to noise ratio = 2.1 and Z´=0.82 was obtained. As a result 49 compounds with PI ≥ 75% were selected for potency screen (Table 3.10).

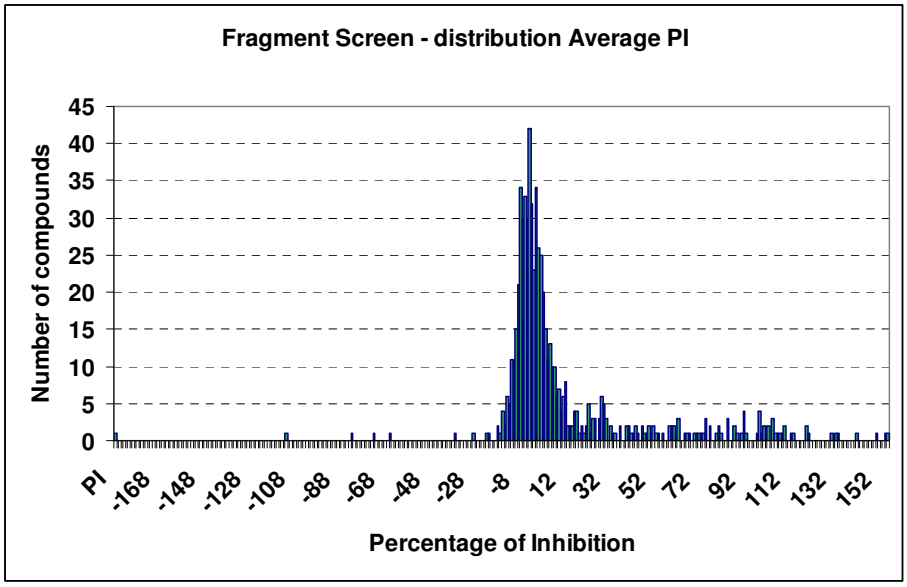


Figure 3.23 PI of fragment set
Distribution of average PI of the fragment set with PI on the x – axis and number of compounds on the y- axis.

The table shows all 49 compounds with the corresponded pIC_{50} ranges, the number of compounds, the number of compounds which inhibited pyrophosphatase and the amount of these compounds which were confirmed as binder in the NMR experiments. Some of the interfering compounds were found to bind to *Tb*UGP in the NMR experiment, which is shown in the red brackets (Table 3.10.)

Table 3.10 Fragment screen result summary

pIC_{50}	# *Tb*UGP	# Interfering with PPase	# *Tb*UGP NMR positives
< 5	6	2	3 (1 of interfering)
5 – 4.7	6	1	4
4-7 – 4.5	4	2	1 (1 of interfering)
4.5 - 4.4	2	-	2
4.4 - 4.3	3	2	1 (1 of interfering)
4.2 - 4	7	2	3
4 – 3.7	13	6	4 (2 of interfering)
3.7 – 3.5	5	-	2
3.5 – 3.4	1	1	1 (1 of interfering)
>3.3	2	-	-

3.3 Discussions of the UGP – project

By using a virtual screening approach, 15 compounds were shortlisted for biochemical testing with *Tb*UGP (Table 3.6) out of which one inhibitor (**16**) was identified. This is a very high hit rate, especially considering that only 15 out of 5.2 million commercially available compounds were tested. Virtual screening was again proven to be an efficient method to find inhibitors in the early drug development stage for a target for which no drug-like inhibitors were known at the onset of the study.

The mode of binding of the virtual screening hit **16** was determined using NMR experiments and enzyme kinetics. NMR experiments showed that **16** binds to *Tb*UGP and competes with UTP for the same binding site (3.1.6, 3.2.3). However, by adding 5 mM UTP the signal in the saturation transfer difference (STD) experiment was not fully reduced (Figure 3.13) and in the wLOGSY experiment not inverted (Figure 3.14). This is probably due to the fact that UTP binds only relatively weakly to *Tb*UGP (K_m: 53 µM, 3.1.4). It is likely, that the competition of the compound would have been much more clearly observed in the spectra if a potent inhibitor for *Tb*UGP was available. In addition to the NMR experiments, studies of the binding kinetics were carried out to determine the mode of inhibition of **16**. By increasing the concentration of UTP, the inhibition of *Tb*UGP decreased, indicating that the inhibitor binds competitively into the substrate binding site (Figure 3.17, Figure 3.18).

To further probe the binding mode of **16**, a point mutation was introduced into *Tb*UGP to close the selectivity pocket as found in the structure of the human homolog (Figure 3.8). In the suggested binding mode (Figure 3.19a) the ligand is occupying this pocket. Hence, it was expected that the compound would no longer inhibit the mutated enzyme. Introducing the mutation proved challenging, because the required primers were prone to intramolecular interactions and formation of secondary structures (Figure 3.24). The site-directed mutagenesis was only successful after adding 10 % DMSO to disturb these unwanted interactions. Analysis of the crystal structure of the mutated enzyme (Figure 3.19b) revealed that the selectivity pocket was not entirely closed as expected based on the structure of *h*UGP but far enough to restrict binding of **16** in the model binding mode of the compound in the wild type enzyme (Figure 3.19a). Compound **16** showed no inhibition with mutated *Tb*UGP up to a concentration of 1 mM which is consistent with the suggested binding mode.

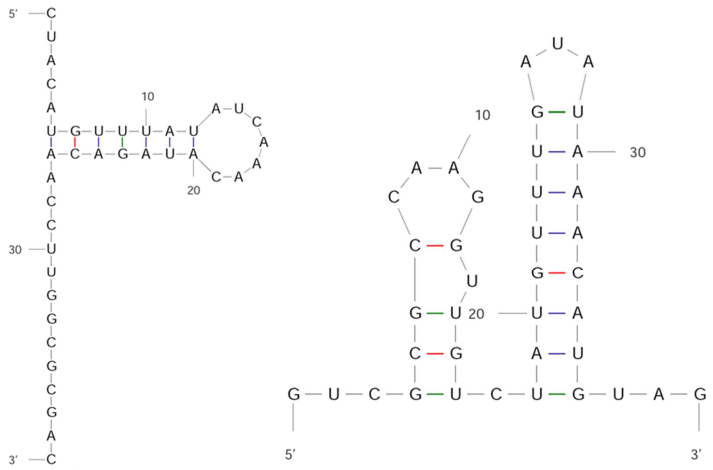


Figure 3.24 Secondary structures of used primers.
Images produced by the sir graph utility of the mfold program (Zuker 2003)

Analogues (Table 3.8) of **16** were purchased and tested for activity to establish SAR. All analogues possess the pyrazole group which is presumably binding into the uracil binding site (Figure 3.19a). They also contain a moiety which can extend into the selectivity cavity. However, none of the compounds have a hydrogen-bond acceptor functionality which corresponds to the nitrogen atom of the bicyclic ring system of **16**, which forms a hydrogen-bond with Gly84. (Note, that compound **23** is similar to **16** but has got an oxygen atom instead of the nitrogen in the bicyclic ring system, however, due to its environment in the ligand it is at best a very poor hydrogen-bond acceptor). Therefore the analogues were expected to be less potent than **16**. All analogues except **26** showed competitive binding in the NMR experiments at 1 mM. It is likely that the nitrogen atom of the pyrazole ring in **26** has no hydrogen-bonding partner in the protein leading to reduced affinity.

Although binding was measured by NMR for the remaining compounds, binding was too weak to be detected in the biochemical assay up to 1 mM. This might indicate that the interaction of the nitrogen atom of the thiazole moiety in **16** with Gly84 is crucial for affinity. To develop further SAR and to improve potency, more analogues must be designed and synthesized.

The virtual screening hit **16** was also tested for inhibition with *h*UGP using the BiomolGreen kinetic assay. Despite the assumption that **16** should not inhibit *h*UGP as the selectivity pocket is closed in the crystal structure (Figure 3.11), a pIC_{50} of 5.3 was measured with a Hill slope of 4.2. This Hill slope is too steep for a competitive inhibition and binding does therefore not follow the law of mass action for a single binding site. A Hill slope > 1 can indicate either non equilibrium state, allosteric binding or positive cooperation (Heck 1971). That would mean that the binding of **16** into one active site of the enzyme increases its affinity for the other sites. These modes of inhibition are different from the one in *Tb*UGP. It is therefore not clear where and how compound **16** is binding to *h*UGP. The amount of *h*UGP enzyme available was not sufficient enough to accomplish further necessary experiments.

Inhibitor **16** has a molecular weight of 201 Daltons and a ligand efficiency of 0.4 kcal/mol per non-hydrogen atom and an alogP of 2.7. This makes it a good starting point for further optimization (Hopkins*, et al.,* 2004). It is possible, that some changes in the fused ring system will not only improve the potency of the compound but also make it more selective for *Tb*UGP as the requirements for binding to the alternatives sites in *h*UGP might be different. With modifications on

16 it will be possible to learn more about the inhibition mode in *h*UGP and develop highly potent and selective inhibitors.

HTS of over 70 000 compounds produced thousands of hits but no obvious SAR. Many hits had Hill slopes that were not consistent with competitive inhibition. Even after clustering the hits in classes and testing them using SPR, the hits could not be confirmed. For this reason an alternative inhibition assay was used. This was carried out by Sabine Kuettel using a Dionex HPLC instrument that directly measured the production of the product UDP-Glc (data not shown). She selected the top hits from the HTS, SPR and fragment screening and tested them by using **7** as a control (Table 3.5). She could not confirm that any of the compounds (including **16**) were active. The reason for the unexpected behaviour of *Tb*UGP under different assay conditions is currently unclear.

4 Summary and Conclusions

In this study the process from assay development to hit validation for two different targets for HAT, *Tb*6PGDH and *Tb*UGP, was carried out. For hit identification different modern techniques were used like *in silico* virtual screening or experimental screening using SPR, NMR and HTS.

4.1 Summary and conclusions regarding 6PGDH – Project

The aim of this project was to use a thermophilic model system like *Gs*6PGDH instead of the highly instable *Tb*6PGDH and search for binders using a fragment screening by NMR (2.1.2). Because the substrate (6PG) pocket of both enzymes is conserved (Figure 2.8) it was a reasonable assumption that compounds found to bind to the model enzyme should also bind in the same way to the target enzyme. The selectivity over *h*6PGDH could then be achieved by chemically expanding the compounds into the co-factor $NADP^+$ - binding site. Structural analysis suggests that the differences around the co-factor should be sufficient to achieve that goal (Figure 2.10). For the kinetic studies a colorimetric assay was developed measuring the production of NADPH at 340 nm (2.1.3.1). This assay is easy to perform but has several limitations. Compounds that absorb at 340 nm or interact with NADPH cannot be measured and are excluded from further studies.

To validate *Gs*6PGDH as a good model it was planned to test compounds which were found to inhibit *Tb*UGP (Ruda*, et al.,* 2010). This first revealed a discrepancy between the published data and the ones collected for the model enzyme (2.2.2). However, this data does not disproved *Gs*6PGDH as a good model enzyme, because the published data lacks of many important details and is therefore uncertain (2.3.1). The fragment screen resulted in many hits from which only two acids were found to weakly bind to *Gs*6PGDH using calorimetric assay (2.2.5). It seems that most of the NMR - hits were not potent enough to be detected by a biochemical assay.

In Summary this study was unfortunately not able to successfully determining a binding mode of published virtual screening hits of *Tb*6PGDH and has not discovered any new potent, fragment-like inhibitors. The polar binding pocket and the instability of the enzyme suggesting that *Tb*6PGDH is not a suitable target for a structure-based hit discovery approach.

4.2 Summary and conclusions regarding UGP – Project

For the target *Tb*UGP a very first, drug-like inhibitor (**16**) was discovered using *in silico* virtual screening. From an in-house virtual library of 5.2 million commercial available compounds, fifteen were purchased for activity test from which two showed activities in the kinetic assay but only for **16** a dose response curve could be measured. Compound **16** has a molecular weight of 201 Dalton, a ligand efficiency

of 0.4 kcal/mol per non-hydrogen atom, an alogP of 2.7 and inhibits *Tb*UGP with a pIC_{50} of 3.53±0.04 and a Hill slope of 1.1±0.1. NMR and enzyme inhibition data for the compound is consistent with it being a competitive inhibitor (3.2.1.3 and 3.2.1.4). This is the first, not substrate like competitive inhibitor found for *Tb*UGP. The proposed binding mode of **16** was evaluated using mutation studies (3.2.1.6) and crystallography (3.1.9). Compound **16** was expected to bind into the UTP pocket and expand into a cavity present in *Tb*UGP (Figure 3.19). A G219I - mutation of *Tb*UGP results in a blocked cavity, similar as found in *h*UGP (Figure 3.10). As expected **16** was no longer inhibiting the G219I–mutant up to 1 mM but was still inhibiting *h*UGP. The proposed selectivity of **16** could not be confirmed in this study. To establish a SAR around compound **16** a substructure search was carried out from which four analogues were purchased (Table 3.8). The NMR – screen showed that three of four compounds bind competitively to *Tb*UGP. A biochemical test showed that none of these compounds were active up to 1 mM. That means that the biochemical assay is not sensitive enough to detect such weak binders.

A HTS with the DDU in-house fragment library using a biochemical BiomolGreen coupled enzyme assay, SPR and NMR came back with a high unusually hit – rate. The best compounds from each test did not coincide with each other. To be able to explain that behaviour the BiomolGreen assay was replaced by an assay using a Dionex HPLC instrument that directly measured the production of the product UDP-Glc. This replacement did not clarify the results but raised even more questions, as none of the hits (including **16**) could be confirmed in this assay. Therefore a new

strategy should be developed to tie up all the results collected and to explain that strange behaviour observed in this study.

In summary, based on the NMR data and the data collected using the BiomolGreen assay with the wild type and G219I–mutant, compound **16** provides a good lead for further development of *Tb*UGP inhibitors. Derivatives of **16** could result in a more potent and selective inhibitor and become a good lead compound for further development.

5 Outlook

6PGDH has proven to be a difficult target. The fact, that no binding modes inhibitors are know which were not substrate like and that the binding pocket of 6PGH enzyme is challenging to address, together with the instability of *Tb*6PGDH, makes this target not a good candidate for further studies. There are better targets for HAT to concentrate on.

One example is UGP. This thesis showed that this enzyme is accessible to a rational hit discovery approach. The crystal structures of *Tb*UGP and *h* which makes it available for structure – based design.

Further studies should be carried out to explain what the mode of action of **16** with *h*UGP is. A crystal structure of *h*UGP with **16** could help to understand the biochemical data and provide guidance for the synthesis of selective derivatives of **16**. A derivative with a higher binding affinity for *Tb*UGP would prove beyond doubt that **16** is not a false positive.

6 Appendix

6.1 Recipe for auto-induction media

Table 6.1 Autoinduction media

<table>
<tr><th colspan="3">For 1 L autoinduction media</th></tr>
<tr><td rowspan="2">5g Yeast extract</td><td>20ml 50x5052</td><td>50x5052:
250g glycerol
25g glucose
10g α-lactose</td></tr>
<tr><td colspan="2">1ml 1M $MgSO_4$</td></tr>
<tr><td>10g Tryptone</td><td colspan="2"></td></tr>
<tr><td>50ml 20xNPS</td><td colspan="2">20xNPS:
66g $(NH_4)_2SO_4$
136g KH_2PO_4
142g Na_2HPO_4</td></tr>
</table>

6.2 Crystallographic statistic tables

Table 6.2 Crystallographic data and refinement statistics of G219-mutatant ligand complex.

Details of Data Collection	
Space group	*C* 2
Unit cell dimensions: *a, b, c* (Å)	166.2, 78.4, 112.3 α=γ=90 °, β=118 °
Resolution range (Å)	45 -2.5
No. reflections	155335
No. unique reflections	41173
Redundancy[a]	3.8 (3.5)
Completeness[a] (%)	93.4 (92.8)
Wilson B (Å^2)	18.2
<I/σ(I)>	8.8 (2.9)
R_{merge}[b] (%)	12.1 (42.8)
Refinement Statistics	
R_{work}[c] / R_{free}[d] (%)	22.9 / 29.6
Number of: Protein residues Chain A Chain B	 466 459

Water molecules	118
UPG (Chain A + B)	2
Average *B*-factors (Å^2)	
Overall / side chain / main chain	20.7 / 21 / 20.3
UPG (chain A / B)	10.3 / 7.8
r.m.s.d. from ideal values:	
Bond lengths (Å)	0.009
Bond angles (˚)	1.360
Ramachandran plot analysis (%)	
Favourable [e]	876 (95.8)
Outliers [e]	2 (0.2)

[a] Values in parentheses refer to the highest resolution shell (2.64-2.5 Å).
[b] $R_{merge} = \sum h \sum i \, ||(h,i) - < \mathrm{I}(h) > \sum h \sum i \, \mathrm{I}(h,i)$; where $\mathrm{I}(h,i)$ is the intensity of the ith measurement of reflection h and < $\mathrm{I}(h)$> is the mean value of $\mathrm{I}(h,i)$ for all i measurements. [c.] $R_{work} = \sum hkl \, ||F_o| - |F_c|| / \sum |F_o|$, where F_o is the observed structure-factor amplitude and the F_c is the structure-factor amplitude calculated from the model. [d.] R_{free} is the same as R_{work} except calculated with a subset, 5 %, of data that are excluded from refinement calculations. [e.] According to MolProbity (http://molprobity.biochem.duke.edu)

Table 6.3 Crystallographic data and refinement statistics of *Gs*6PGDH.

Details of Data Collection	
Space group	$P2_12_12_1$
Unit cell dimensions: *a, b, c* (Å)	67.8, 119.8, 141.9
Resolution range (Å)	38.8 -2.7
No. reflections	144947
No. unique reflections	31971
Redundancy[a]	4.5 (4.0)
Completeness[a] (%)	98.6 (97.0)
Wilson B (Å^2)	18.2
$\langle I/\sigma(I)\rangle$	11.2 (2.5)
R_{merge}[b] (%)	13.1 (57.6)
Refinement Statistics	
R_{work}[c] / R_{free}[d] (%)	21.2 / 27.9
Number of:	
Protein residues	
Chain A	464
Chain B	466
Water molecules	26
Average *B*-factors (Å^2)	
Overall / side chain / main chain	26.7 / 27.5 / 25.9

r.m.s.d. from ideal values:	
Bond lengths (Å)	0.01
Bond angles (˚)	1.347
Ramachandran plot analysis (%)	
Favourable [e]	889 (96.3)
Outliers [e]	3 (0.3)

[a] Values in parentheses refer to the highest resolution shell (2.64-2.5 Å).
[b] $R_{merge} = \sum h \sum i \, ||(h,i) - < \mathrm{I}(h) > \sum h \sum i \, \mathrm{I}(h,i)$; where $\mathrm{I}(h,i)$ is the intensity of the ith measurement of reflection h and < $\mathrm{I}(h)$> is the mean value of $\mathrm{I}(h,i)$ for all i measurements. [c.] $R_{work} = \sum hkl \, ||F_o| - |F_c|| / \sum |F_o|$, where F_o is the observed structure-factor amplitude and the F_c is the structure-factor amplitude calculated from the model. [d.] R_{free} is the same as R_{work} except calculated with a subset, 5 %, of data that are excluded from refinement calculations. [e.] According to MolProbity (http://molprobity.biochem.duke.edu)

6.3 List of Abbreviations

- °C – degree Celsius
- µ – mycro
- 6PGDH – 6-Phosphogluconate dehydrogenase
- CHAPS – 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate
- DMSO – dimethyl sulfoxide
- DNA – deoxyribonucleic acid
- Dnase – desoxyribonuklease
- DTT – dithiothreitol
- EDTA – ethylenediamine tetraacetic acid
- EGTA – ethylene glycol tetraacetic acid
- ESRF – European synchrotron radiation facility
- G – earth´s gravitational acceleration
- g – Gramm
- *Gs* – *Geobacillus stearothermophilus*
- h – hour/s
- *h* - human
- HEPES – 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
- HPLC – high performance liquid chromatography
- IPTG – Isopropyl-β-D-thiogalactopyranosid
- K – Kelvin
- kDa - kilodalton
- kg – Kilogramm
- l – litre
- m – milli
- M – Molar
- $NADP^+$ – nicotinamide adenine dinucleotide phosphate (reduced form NADPH)
- nm – nanometre
- o/n – over night
- OD_{600} optical density at 600nm
- ORF – open reading frame
- PCR – Polymerase chain reaction
- PDB – protein databank
- PEG – polyethylene glycol
- PPase -pyrophosphatase

- QSAR – quantitavie structure-activity relationship
- RMS – root mean square
- RMSD – root mean square deviation
- Rpm – rounds per minute
- TB – Terrific Broth (media)
- *Tb* – *Trypanosoma brucei*
- TEV – Tobacco Etch Virus
- Tris – 2-amino-hydroxymethyl-propane-1,3-diol
- U – Uracil
- UDP – Uridine diphosphate
- UGP – UDP-glucose pyrophosphorylase

7 References

Abreu, R. M., H. J. Froufe, M. J. Queiroz and I. C. Ferreira (2012). "Selective flexibility of side-chain residues improves VEGFR-2 docking score using AutoDock Vina." Chem Biol Drug Des **79**(4): 530-534.

Alphey, M. S., L. Pirrie, L. S. Torrie, W. A. Boulkeroura, M. Gardiner, A. Sarkar, M. Maringer, W. Oehlmann, R. Brenk, M. S. Scherman, M. McNeil, M. Rejzek, R. A. Field, M. Singh, D. Gray, N. J. Westwood and J. H. Naismith (2012). "Allosteric Competitive Inhibitors of the Glucose-1-Phosphate Thymidylyltransferase (RmlA) from Pseudomonas aeruginosa." ACS Chem Biol.

Arnold, K., L. Bordoli, J. Kopp and T. Schwede (2006). "The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling." Bioinformatics **22**(2): 195-201.

Aronov, A. M., S. Suresh, F. S. Buckner, W. C. Van Voorhis, C. L. Verlinde, F. R. Opperdoes, W. G. Hol and M. H. Gelb (1999). "Structure-based design of submicromolar, biologically active inhibitors of trypanosomatid glyceraldehyde-3-phosphate dehydrogenase." Proc Natl Acad Sci U S A **96**(8): 4273-4278.

Artimo, P., M. Jonnalagedda, K. Arnold, D. Baratin, G. Csardi, E. de Castro, S. Duvaud, V. Flegel, A. Fortier, E. Gasteiger, A. Grosdidier, C. Hernandez, V. Ioannidis, D. Kuznetsov, R. Liechti, S. Moretti, K. Mostaguir, N. Redaschi, G. Rossier, I. Xenarios and H. Stockinger (2012). "ExPASy: SIB bioinformatics resource portal." Nucleic Acids Res **40**(Web Server issue): W597-603.

Baba, Y., N. Hirukawa, N. Tanohira and M. Sodeoka (2003). "Structure-based design of a highly selective catalytic site-directed inhibitor of Ser/Thr protein phosphatase 2B (calcineurin)." J Am Chem Soc **125**(32): 9740-9749.

Badger, J. (2012). "Crystallographic fragment screening." Methods Mol Biol **841**: 161-177.

Balasegaram, M., S. Balasegaram, D. Malvy and P. Millet (2008). "Neglected diseases in the news: a content analysis of recent international media coverage focussing on leishmaniasis and trypanosomiasis." PLoS Negl Trop Dis **2**(5): e234.

Barrett, M. P. and R. W. Le Page (1993). "A 6-phosphogluconate dehydrogenase gene from Trypanosoma brucei." Mol Biochem Parasitol **57**(1): 89-99.

Bastin, P., T. Sherwin and K. Gull (1998). "Paraflagellar rod is vital for trypanosome motility." Nature **391**(6667): 548.

Battye, T. G., L. Kontogiannis, O. Johnson, H. R. Powell and A. G. Leslie (2011). "iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM." Acta Crystallogr D Biol Crystallogr **67**(Pt 4): 271-281.

Bender, A., J. L. Jenkins, J. Scheiber, S. C. Sukuru, M. Glick and J. W. Davies (2009). "How similar are similarity searching methods? A principal component analysis of molecular descriptor space." J Chem Inf Model **49**(1): 108-119.

Bernardo, P. H. and J. C. Tong (2012). "In silico design of small molecules." Methods Mol Biol **800**: 25-31.

Bertelli, M., E. El-Bastawissy, M. H. Knaggs, M. P. Barrett, S. Hanau and I. H. Gilbert (2001). "Selective inhibition of 6-phosphogluconate dehydrogenase from Trypanosoma brucei." J Comput Aided Mol Des **15**(5): 465-475.

Bhunia, A., S. Bhattacharjya and S. Chatterjee (2012). "Applications of saturation transfer difference NMR in biological systems." Drug Discov Today **17**(9-10): 505-513.

Bickel, M. H. (1988). "The development of sulfonamides (1932-1938) as a focal point in the history of chemotherapy." Gesnerus **45 Pt 1**: 67-86.

Bocker, A., P. R. Bonneau and P. J. Edwards (2011). "HTS promiscuity analyses for accelerating decision making." J Biomol Screen **16**(7): 765-774.

Bohacek, R. S., C. McMartin and W. C. Guida (1996). "The art and practice of structure-based drug design: a molecular modeling perspective." Med Res Rev **16**(1): 3-50.

Borst, P., G. Rudenko, M. C. Taylor, P. A. Blundell, F. Van Leeuwen, W. Bitter, M. Cross and R. McCulloch (1996). "Antigenic variation in trypanosomes." Arch Med Res **27**(3): 379-388.

Bower, J. F. and A. Pannifer (2012). "Using fragment-based technologies to target protein-protein interactions." Curr Pharm Des **18**(30): 4685-4696.

Brenk, R., A. Schipani, D. James, A. Krasowski, I. H. Gilbert, J. Frearson and P. G. Wyatt (2008). "Lessons learnt from assembling screening libraries for drug discovery for neglected diseases." ChemMedChem **3**(3): 435-444.

Brun, R., J. Blum, F. Chappuis and C. Burri (2010). "Human African trypanosomiasis." Lancet **375**(9709): 148-159.

Cameron, S., V. P. Martini, J. Iulek and W. N. Hunter (2009). "Geobacillus stearothermophilus 6-phosphogluconate dehydrogenase complexed with 6-phosphogluconate." Acta Crystallogr Sect F Struct Biol Cryst Commun **65**(Pt 5): 450-454.

Chatelain, E. and J. R. Ioset (2011). "Drug discovery and development for neglected diseases: the DNDi model." Drug Des Devel Ther **5**: 175-181.

Checchi, F. and M. P. Barrett (2008). "African sleeping sickness." BMJ **336**(7646): 679-680.

Chen, V. B., W. B. Arendall, 3rd, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson and D. C. Richardson (2010). "MolProbity: all-atom structure validation for macromolecular crystallography." Acta Crystallogr D Biol Crystallogr **66**(Pt 1): 12-21.

Cheng, T., Q. Li, Z. Zhou, Y. Wang and S. H. Bryant (2012). "Structure-based virtual screening for drug discovery: a problem-centric review." AAPS J **14**(1): 133-141.

Cheng, T. O. (2007). "The history of aspirin." Tex Heart Inst J **34**(3): 392-393.

Cheng, Y. and W. H. Prusoff (1973). "Relationship between the inhibition constant (K1) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction." Biochem Pharmacol **22**(23): 3099-3108.

Cleghorn, L. A., A. Woodland, I. T. Collie, L. S. Torrie, N. Norcross, T. Luksch, C. Mpamhanga, R. G. Walker, J. C. Mottram, R. Brenk, J. A. Frearson, I. H. Gilbert and P. G. Wyatt (2011). "Identification of inhibitors of the Leishmania cdc2-related protein kinase CRK3." ChemMedChem **6**(12): 2214-2224.

Coupez, B. and R. A. Lewis (2006). "Docking and scoring--theoretically easy, practically impossible?" Curr Med Chem **13**(25): 2995-3003.

Courtin, F., V. Jamonneau, G. Duvallet, M. Camara, D. Kaba and P. Solano (2008). "[One century of "sleeping sickness" in West Africa]." Bull Soc Pathol Exot **101**(3): 287-289.

Craig, M. L., O. V. Tsodikov, K. L. McQuade, P. E. Schlax, Jr., M. W. Capp, R. M. Saecker and M. T. Record, Jr. (1998). "DNA footprints of the two kinetically significant intermediates in formation of an RNA polymerase-promoter open complex: evidence that interactions with start site and downstream DNA induce sequential conformational changes in polymerase and DNA." J Mol Biol **283**(4): 741-756.

Crisman, T. J., C. N. Parker, J. L. Jenkins, J. Scheiber, M. Thoma, Z. B. Kang, R. Kim, A. Bender, J. H. Nettles, J. W. Davies and M. Glick (2007). "Understanding false positives in reporter gene assays: in silico chemogenomics approaches to prioritize cell-based HTS data." J Chem Inf Model **47**(4): 1319-1327.

Cronin, C. N., D. P. Nolan and H. P. Voorheis (1989). "The enzymes of the classical pentose phosphate pathway display differential activities in procyclic and bloodstream forms of Trypanosoma brucei." FEBS Lett **244**(1): 26-30.

Cross, G. A. (1996). "Antigenic variation in trypanosomes: secrets surface slowly." Bioessays **18**(4): 283-291.

Dalvit, C. (2009). "NMR methods in fragment screening: theory and a comparison with other biophysical techniques." Drug Discov Today **14**(21-22): 1051-1057.

Dalvit, C. and J. Böhlen (1996). "Simultaneous Suppression of the H2O Double-Quantum Signal and of the Radiation-Damping Effect in Double-Quantum Experiments." J Magn Reson B **113**(2): 195-200.

Dalvit, C. and J. Böhlen (1996). "Spectral Editing and Water Suppression in Double-Quantum Experiments." J Magn Reson B **111**(1): 76-80.

Dalvit, C., P. Pevarello, M. Tato, M. Veronesi, A. Vulpetti and M. Sundstrom (2000). "Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water." J Biomol NMR **18**(1): 65-68.

Dardonville, C., E. Rinaldi, M. P. Barrett, R. Brun, I. H. Gilbert and S. Hanau (2004). "Selective inhibition of Trypanosoma brucei 6-phosphogluconate dehydrogenase by high-energy intermediate and transition-state analogues." J Med Chem **47**(13): 3427-3437.

Duong-Thi, M. D., M. Bergstrom, T. Fex, R. Isaksson and S. Ohlson (2012). "High-Throughput Fragment Screening by Affinity LC-MS." J Biomol Screen.

Duschak, V. G. (2011). "A decade of targets and patented drugs for chemotherapy of Chagas disease." Recent Pat Antiinfect Drug Discov **6**(3): 216-259.

Ekins, S., J. Mestres and B. Testa (2007). "In silico pharmacology for drug discovery: applications to targets and beyond." Br J Pharmacol **152**(1): 21-37.

Ekins, S., J. Mestres and B. Testa (2007). "In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling." Br J Pharmacol **152**(1): 9-20.

Elebring, T., A. Gill and A. T. Plowright (2012). "What is the most important approach in current drug discovery: doing the right things or doing things right?" Drug Discov Today **17**(21-22): 1166-1169.

Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2126-2132.

Emsley, P., B. Lohkamp, W. G. Scott and K. Cowtan (2010). "Features and development of Coot." Acta Crystallogr D Biol Crystallogr **66**(Pt 4): 486-501.

Evans, P. R. (2006). "Scaling and assessment of data quality." Acta Crystallogr D Biol Crystallogr **62**(Pt 1): 72-82.

Evans, P. R. (2011). "An introduction to data reduction: space-group determination, scaling and intensity statistics." Acta Crystallogr D Biol Crystallogr **67**(Pt 4): 282-292.

Feher, M. and C. I. Williams (2010). "Reducing docking score variations arising from input differences." J Chem Inf Model **50**(9): 1549-1560.

Fevre, E. M., B. V. Wissmann, S. C. Welburn and P. Lutumba (2008). "The burden of human African trypanosomiasis." PLoS Negl Trop Dis **2**(12): e333.

Fink, T., H. Bruggesser and J. L. Reymond (2005). "Virtual exploration of the small-molecule chemical universe below 160 Daltons." Angew Chem Int Ed Engl **44**(10): 1504-1508.

Fire, A., S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver and C. C. Mello (1998). "Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans." Nature **391**(6669): 806-811.

Flores-Diaz, M., A. Alape-Giron, B. Persson, P. Pollesello, M. Moos, C. von Eichel-Streiber, M. Thelestam and I. Florin (1997). "Cellular UDP-glucose deficiency caused by a single point mutation in the UDP-glucose pyrophosphorylase gene." J Biol Chem **272**(38): 23784-23791.

Folts, J. D. (2007). "The history of aspirin." Tex Heart Inst J **34**(3): 392.

Friesner, R. A., J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis and P. S. Shenkin (2004). "Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy." J Med Chem **47**(7): 1739-1749.

Gerber, P. R. and K. Muller (1995). "MAB, a generally applicable molecular force field for structure modelling in medicinal chemistry." J Comput Aided Mol Des **9**(3): 251-268.

Gouet, P., X. Robert and E. Courcelle (2003). "ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins." Nucleic Acids Res **31**(13): 3320-3323.

Graves, A. P., R. Brenk and B. K. Shoichet (2005). "Decoys for docking." J Med Chem **48**(11): 3714-3728.

Halgren, T. A., R. B. Murphy, R. A. Friesner, H. S. Beard, L. L. Frye, W. T. Pollard and J. L. Banks (2004). "Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening." J Med Chem **47**(7): 1750-1759.

Hammond, C. and A. Helenius (1995). "Quality control in the secretory pathway." Curr Opin Cell Biol **7**(4): 523-529.

Hanau, S., E. Rinaldi, F. Dallocchio, I. H. Gilbert, C. Dardonville, M. J. Adams, S. Gover and M. P. Barrett (2004). "6-phosphogluconate dehydrogenase: a target for drugs in African trypanosomes." Curr Med Chem **11**(19): 2639-2650.

Hanau, S., M. Rippa, M. Bertelli, F. Dallocchio and M. P. Barrett (1996). "6-Phosphogluconate dehydrogenase from Trypanosoma brucei. Kinetic analysis and inhibition by trypanocidal drugs." Eur J Biochem **240**(3): 592-599.

Hann, M. M., A. R. Leach and G. Harper (2001). "Molecular complexity and its impact on the probability of finding leads for drug discovery." J Chem Inf Comput Sci **41**(3): 856-864.

Hawkins, P. C. D., A. G. Skillman, G. L. Warren, B. A. Ellingson and M. T. Stahl (2010). "Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database." Journal of Chemical Information and Modeling **50**(4): 572-584.

Heck, H. D. (1971). "Statistical theory of cooperative binding to proteins. The Hill equation and the binding potential." J Am Chem Soc **93**(1): 23-29.

Hopkins, A. L., C. R. Groom and A. Alex (2004). "Ligand efficiency: a useful metric for lead selection." Drug Discov Today **9**(10): 430-431.

Hurko, O. (2012). "Target-based drug discovery, genetic diseases, and biologics." Neurochem Int **61**(6): 892-898.

Jacobs, R. T., B. Nare and M. A. Phillips (2011). "State of the art in African trypanosome drug discovery." Curr Top Med Chem **11**(10): 1255-1274.

Jones, G., P. Willett, R. C. Glen, A. R. Leach and R. Taylor (1997). "Development and validation of a genetic algorithm for flexible docking." J Mol Biol **267**(3): 727-748.

Jordan, J. B., L. Poppe, X. Xia, A. C. Cheng, Y. Sun, K. Michelsen, H. Eastwood, P. D. Schnier, T. Nixey and W. Zhong (2012). "Fragment based drug discovery: practical implementation based on (1)(9)F NMR spectroscopy." J Med Chem **55**(2): 678-687.

Kappagoda, S., U. Singh and B. G. Blackburn (2011). "Antiparasitic therapy." Mayo Clin Proc **86**(6): 561-583.

Karsten, W. E., L. Chooback and P. F. Cook (1998). "Glutamate 190 is a general acid catalyst in the 6-phosphogluconate-dehydrogenase-catalyzed reaction." Biochemistry **37**(45): 15691-15697.

Knowles, J. and G. Gromo (2003). "A guide to drug discovery: Target selection in drug discovery." Nat Rev Drug Discov **2**(1): 63-69.

Kramer, B., M. Rarey and T. Lengauer (1999). "Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking." Proteins **37**(2): 228-241.

Kumar, A., A. Voet and K. Y. Zhang (2012). "Fragment based drug design: from experimental to computational approaches." Curr Med Chem.

Kuntz, I. D., J. M. Blaney, S. J. Oatley, R. Langridge and T. E. Ferrin (1982). "A geometric approach to macromolecule-ligand interactions." J Mol Biol **161**(2): 269-288.

Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson and D. G. Higgins (2007). "Clustal W and Clustal X version 2.0." Bioinformatics **23**(21): 2947-2948.

Leroux, A. E., D. A. Maugeri, F. R. Opperdoes, J. J. Cazzulo and C. Nowicki (2011). "Comparative studies on the biochemical properties of the malic enzymes from Trypanosoma cruzi and Trypanosoma brucei." FEMS Microbiol Lett **314**(1): 25-33.

Leslie, A. G. W. and H. R. Powell (2007). "Evolving Methods for Macromolecular Crystallography." NATO Series II **vol 245**: 41–51.

Lindesmith, A. R. (1968). "Addiction and Opiates." 295.

Lipinski, C. A., F. Lombardo, B. W. Dominy and P. J. Feeney (1997). "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings." Adv Drug Deliv Rev **23**(1-3): 3-25.

Lipinski, C. A., F. Lombardo, B. W. Dominy and P. J. Feeney (2001). "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings." Adv Drug Deliv Rev **46**(1-3): 3-26.

Liu, Y., M. H. Beresini, A. Johnson, R. Mintzer, K. Shah, K. Clark, S. Schmidt, C. Lewis, M. Liimatta, L. O. Elliott, A. Gustafson and C. E. Heise (2012). "Case studies of minimizing nonspecific inhibitors in HTS campaigns that use assay-ready plates." J Biomol Screen **17**(2): 225-236.

Lorber, D. M. and B. K. Shoichet (1998). "Flexible ligand docking using conformational ensembles." Protein Sci **7**(4): 938-950.

Lorber, D. M. and B. K. Shoichet (2005). "Hierarchical docking of databases of multiple ligand conformations." Curr Top Med Chem **5**(8): 739-749.

Ma, X. H., F. Zhu, X. Liu, Z. Shi, J. X. Zhang, S. Y. Yang, Y. Q. Wei and Y. Z. Chen (2012). "Virtual screening methods as tools for drug lead discovery from large chemical libraries." Curr Med Chem.

Malvy, D. and F. Chappuis (2011). "Sleeping sickness." Clin Microbiol Infect **17**(7): 986-995.

Marchand, M., U. Kooystra, R. K. Wierenga, A. M. Lambeir, J. Van Beeumen, F. R. Opperdoes and P. A. Michels (1989). "Glucosephosphate isomerase from Trypanosoma brucei. Cloning and characterization of the gene and analysis of the enzyme." Eur J Biochem **184**(2): 455-464.

Marino, K., M. L. Guther, A. K. Wernimont, W. Qiu, R. Hui and M. A. Ferguson (2011). "Characterization, localization, essentiality, and high-resolution crystal structure of glucosamine 6-phosphate N-acetyltransferase from Trypanosoma brucei." Eukaryot Cell **10**(7): 985-997.

Marino, K., M. L. S. Guther, A. K. Wernimont, M. Amani, R. Hui and M. A. J. Ferguson (2010). "Identification, subcellular localization, biochemical properties, and high-resolution crystal structure of Trypanosoma brucei UDP-glucose pyrophosphorylase." Glycobiology **20**(12): 1619-1630.

Matter, H., K. H. Baringhaus, T. Naumann, T. Klabunde and B. Pirard (2001). "Computational approaches towards the rational design of drug-like compound libraries." Comb Chem High Throughput Screen **4**(6): 453-475.

Maugeri, D. A., J. J. Cannata and J. J. Cazzulo (2011). "Glucose metabolism in Trypanosoma cruzi." Essays Biochem **51**: 15-30.

McCoy, A. J. (2007). "Solving structures of protein complexes by molecular replacement with Phaser." Acta Crystallogr D Biol Crystallogr **63**(Pt 1): 32-41.

McCoy, A. J., R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read (2007). "Phaser crystallographic software." J Appl Crystallogr **40**(Pt 4): 658-674.

McGann, M. R., H. R. Almond, A. Nicholls, J. A. Grant and F. K. Brown (2003). "Gaussian docking functions." Biopolymers **68**(1): 76-90.

Meng, E. C., B. K. Shoichet and I. D. Kuntz (1992). "Atomated docking with grid-based energy evaluation." J Comput Chem **13**: 505-524.

Meyer, M. M. B. (1999). "Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy." Angewandte Chemie International Edition **38**(12): 1784-1788.

Meyer, M. M. B. (1999). "A Fast and Sensitive Method to Characterize Ligand Binding by Saturation Transfer Difference NMR Spectra,." Angewandte Chemie **38**(111): 1902-1906

Migchelsen, S. J., P. Buscher, A. I. Hoepelman, H. D. Schallig and E. R. Adams (2011). "Human African trypanosomiasis: a review of non-endemic cases in the past 20 years." Int J Infect Dis **15**(8): e517-524.

Mizukoshi, Y., A. Abe, T. Takizawa, H. Hanzawa, Y. Fukunishi, I. Shimada and H. Takahashi (2012). "An accurate pharmacophore mapping method by NMR spectroscopy." Angew Chem Int Ed Engl **51**(6): 1362-1365.

Murshudov, G. N., P. Skubak, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long and A. A. Vagin (2011). "REFMAC5 for the refinement of

macromolecular crystal structures." Acta Crystallogr D Biol Crystallogr **67**(Pt 4): 355-367.

Mysinger, M. M. and B. K. Shoichet (2010). "Rapid Context-Dependent Ligand Desolvation in Molecular Docking." Journal of Chemical Information and Modeling **50**(9): 1561-1573.

Nicholls, A. and B. Honig (1991). "A Rapid Finite-Difference Algorithm, Utilizing Successive over-Relaxation to Solve the Poisson-Boltzmann Equation." J Comput Chem **12**(4): 435-445.

Oprea, T. I., A. M. Davis, S. J. Teague and P. D. Leeson (2001). "Is there a difference between leads and drugs? A historical perspective." J Chem Inf Comput Sci **41**(5): 1308-1315.

Otwinowski, Z. and W. Minor (1997). "Processing of X-ray Diffraction Data Collected in Oscillation Mode." Macromolecular Crystallography **vol 276**: 307–326.

Pasti, C., E. Rinaldi, C. Cervellati, F. Dallocchio, R. Hardre, L. Salmon and S. Hanau (2003). "Sugar derivatives as new 6-phosphogluconate dehydrogenase inhibitors selective for the parasite Trypanosoma brucei." Bioorg Med Chem **11**(7): 1207-1214.

Pencheva, T., O. S. Soumana, I. Pajeva and M. A. Miteva (2010). "Post-docking virtual screening of diverse binding pockets: comparative study using DOCK, AMMOS, X-Score and FRED scoring functions." Eur J Med Chem **45**(6): 2622-2628.

Phillips, C., J. Dohnalek, S. Gover, M. P. Barrett and M. J. Adams (1998). "A 2.8 A resolution structure of 6-phosphogluconate dehydrogenase from the protozoan parasite Trypanosoma brucei: comparison with the sheep enzyme accounts for differences in activity with coenzyme and substrate analogues." J Mol Biol **282**(3): 667-681.

Posner, B. A., H. Xi and J. E. Mills (2009). "Enhanced HTS hit selection via a local hit rate analysis." J Chem Inf Model **49**(10): 2202-2210.

Potterton, E., P. Briggs, M. Turkenburg and E. Dodson (2003). "A graphical user interface to the CCP4 program suite." Acta Crystallogr D Biol Crystallogr **59**(Pt 7): 1131-1137.

Prummer, M. (2012). "Hypothesis testing in high-throughput screening for drug discovery." J Biomol Screen **17**(4): 519-529.

Roper, J. R., M. L. Guther, J. I. Macrae, A. R. Prescott, I. Hallyburton, A. Acosta-Serrano and M. A. Ferguson (2005). "The suppression of galactose metabolism in procylic form Trypanosoma brucei causes cessation of cell growth and alters procyclin glycoprotein structure and copy number." J Biol Chem **280**(20): 19728-19736.

Roper, J. R., M. L. Guther, K. G. Milne and M. A. Ferguson (2002). "Galactose metabolism is essential for the African sleeping sickness parasite Trypanosoma brucei." Proc Natl Acad Sci U S A **99**(9): 5884-5889.

Rovere, F. and G. Gastaldi (1967). "[Study of the Embden Meyerhof Parnas glycolytic pathway in different physiopathologic states of the human endometrium]." Pathologica **59**(871): 83-86.

Ruda, G. F., V. P. Alibu, C. Mitsos, O. Bidet, M. Kaiser, R. Brun, M. P. Barrett and I. H. Gilbert (2007). "Synthesis and biological evaluation of phosphate prodrugs of 4-phospho-D-erythronohydroxamic acid, an inhibitor of 6-phosphogluconate dehydrogenase." ChemMedChem **2**(8): 1169-1180.

Ruda, G. F., G. Campbell, V. P. Alibu, M. P. Barrett, R. Brenk and I. H. Gilbert (2010). "Virtual fragment screening for novel inhibitors of 6-phosphogluconate dehydrogenase." Bioorg Med Chem **18**(14): 5056-5062.

Ruda, G. F., P. E. Wong, V. P. Alibu, S. Norval, K. D. Read, M. P. Barrett and I. H. Gilbert (2010). "Aryl phosphoramidates of 5-phospho erythronohydroxamic acid, a new class of potent trypanocidal compounds." J Med Chem **53**(16): 6071-6078.

Sali, A. and T. L. Blundell (1993). "Comparative protein modelling by satisfaction of spatial restraints." J Mol Biol **234**(3): 779-815.

Schirren, C. (1988). "[The history of dermatology. "Sulfonamides and penicillins"--in memory of the appearance of this book and in memory of Walther Schonfeld, who was born 100 years ago]." Z Hautkr **63**(11): 890.

Schror, K. (2009). "[100 years of successful drug discovery. The history of aspirin]." Pharm Unserer Zeit **38**(4): 306-313.

Sheridan, R. P. and S. K. Kearsley (2002). "Why do we need so many chemical similarity search methods?" Drug Discov Today **7**(17): 903-911.

Shoichet, B. K., D. L. Bodian and I. D. Kuntz (1992). "Molecular docking using shape descriptors." J Comput Chem **13**: 380-397.

Shun, T. Y., J. S. Lazo, E. R. Sharlow and P. A. Johnston (2011). "Identifying actives from HTS data sets: practical approaches for the selection of an appropriate HTS data-processing method and quality control review." J Biomol Screen **16**(1): 1-14.

Simarro, P. P., A. Diarra, J. A. Ruiz Postigo, J. R. Franco and J. G. Jannin (2011). "The human African trypanosomiasis control and surveillance programme of the World Health Organization 2000-2009: the way forward." PLoS Negl Trop Dis **5**(2): e1007.

Simarro, P. P., J. R. Franco, G. Cecchi, M. Paone, A. Diarra, J. A. Ruiz Postigo and J. G. Jannin (2012). "Human African trypanosomiasis in non-endemic countries (2000-2010)." J Travel Med **19**(1): 44-53.

Sink, R., S. Gobec, S. Pecar and A. Zega (2010). "False positives in the early stages of drug discovery." Curr Med Chem **17**(34): 4231-4255.

Sneader, W. (2000). "The discovery of aspirin: a reappraisal." BMJ **321**(7276): 1591-1594.

Stark, J. L. and R. Powers (2012). "Application of NMR and molecular docking in structure-based drug discovery." Top Curr Chem **326**: 1-34.

Stern, A. L., A. Naworyta, J. J. Cazzulo and S. L. Mowbray (2011). "Structures of type B ribose 5-phosphate isomerase from Trypanosoma cruzi shed light on the determinants of sugar specificity in the structural family." Febs J **278**(5): 793-808.

Stich, A., P. M. Abel and S. Krishna (2002). "Human African trypanosomiasis." BMJ **325**(7357): 203-206.

Stokes, M. J., M. L. Guther, D. C. Turnock, A. R. Prescott, K. L. Martin, M. S. Alphey and M. A. Ferguson (2008). "The synthesis of UDP-N-acetylglucosamine is essential for bloodstream form trypanosoma brucei in vitro and in vivo and UDP-N-acetylglucosamine starvation reveals a hierarchy in parasite protein glycosylation." J Biol Chem **283**(23): 16147-16161.

Sundaramoorthy, R., J. Iulek, M. P. Barrett, O. Bidet, G. F. Ruda, I. H. Gilbert and W. N. Hunter (2007). "Crystal structures of a bacterial 6-phosphogluconate dehydrogenase reveal aspects of specificity, mechanism and mode of inhibition by analogues of high-energy reaction intermediates." Febs J **274**(1): 275-286.

T.L. Hwang, A. J. S. (1995). "Water suppression that works. Excitation sculpting using arbitrary waveforms and pulsed field gradients." J. Magn. Reson.(112): 275–279.

Teng, M., M. T. Hilgers, M. L. Cunningham, A. Borchardt, J. B. Locke, S. Abraham, G. Haley, B. P. Kwan, C. Hall, G. W. Hough, K. J. Shaw and J. Finn (2013). "Identification of bacteria-selective threonyl-tRNA synthetase substrate inhibitors by structure-based design." J Med Chem **56**(4): 1748-1760.

Thompson, J. D., D. G. Higgins and T. J. Gibson (1994). "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." Nucleic Acids Res **22**(22): 4673-4680.

Turnock, D. C. and M. A. Ferguson (2007). "Sugar nucleotide pools of Trypanosoma brucei, Trypanosoma cruzi, and Leishmania major." Eukaryot Cell **6**(8): 1450-1463.

Turnock, D. C., L. Izquierdo and M. A. Ferguson (2007). "The de novo synthesis of GDP-fucose is essential for flagellar adhesion and cell growth in Trypanosoma brucei." J Biol Chem **282**(39): 28853-28863.

Urbaniak, M. D., D. C. Turnock and M. A. Ferguson (2006). "Galactose starvation in a bloodstream form Trypanosoma brucei UDP-glucose 4'-epimerase conditional null mutant." Eukaryot Cell **5**(11): 1906-1913.

van Leeuwen, F., R. Kieft, M. Cross and P. Borst (1998). "Biosynthesis and function of the modified DNA base beta-D-glucosyl-hydroxymethyluracil in Trypanosoma brucei." Mol Cell Biol **18**(10): 5643-5651.

van Leeuwen, F., E. R. Wijsman, R. Kieft, G. A. van der Marel, J. H. van Boom and P. Borst (1997). "Localization of the modified base J in telomeric VSG gene expression sites of Trypanosoma brucei." Genes Dev **11**(23): 3232-3241.

van Miert, A. S. (1994). "The sulfonamide-diaminopyrimidine story." J Vet Pharmacol Ther **17**(4): 309-316.

Venkatachalam, C. M., X. Jiang, T. Oldfield and M. Waldman (2003). "LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites." J Mol Graph Model **21**(4): 289-307.

Vilar, S., R. Harpaz, E. Uriarte, L. Santana, R. Rabadan and C. Friedman (2012). "Drug-drug interaction through molecular structure similarity analysis." J Am Med Inform Assoc **19**(6): 1066-1074.

Wang, J., Cieplak, P., and Kollman, P. A. (2000). "How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?" Journal of Computational Chemistry **21**(12): 1049–1074.

Wang, J. and S. Li (2006). "Catalytic mechanism of 6-phosphogluconate dehydrogenase: a theoretical investigation." J Phys Chem B **110**(13): 7029-7035.

Wei, B. Q., W. A. Baase, L. H. Weaver, B. W. Matthews and B. K. Shoichet (2002). "A model binding site for testing scoring functions in molecular docking." J Mol Biol **322**(2): 339-355.

Weininger, D. (1988). "SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules." J Chem Inf Comput Sci **28**(1): 31-36.

Wick, J. Y. (2012). "Aspirin: a history, a love story." Consult Pharm **27**(5): 322-329.

Wirmer-Bartoschek, J. and S. Bartoschek (2012). "NMR in drug discovery on membrane proteins." Future Med Chem **4**(7): 869-875.

Yu, Q. and X. Zheng (2012). "The crystal structure of human UDP-glucose pyrophosphorylase reveals a latch effect that influences enzymatic activity." Biochem J **442**(2): 283-291.

Zhang, J. H., T. D. Chung and K. R. Oldenburg (1999). "A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays." J Biomol Screen **4**(2): 67-73.

Zhang, L., L. Chooback and P. F. Cook (1999). "Lysine 183 is the general base in the 6-phosphogluconate dehydrogenase-catalyzed reaction." Biochemistry **38**(35): 11231-11238.

Zuker, M. (2003). "Mfold web server for nucleic acid folding and hybridization prediction." Nucleic Acids Res **31**(13): 3406-3415.